11th Edition

Automotive Handbook



BOSCH
Invented for life

WILEY



BOSCH

Invented for life

Automotive Handbook 11th Edition



BOSCH

Invented for life

Automotive Handbook

Imprint

Published by:
© Robert Bosch GmbH
Postfach 410960
D-76225 Karlsruhe
Business Division
Automotive Aftermarket

Scientific advisor and editor
Prof. Dr.-Ing. Konrad Reif
Duale Hochschule Baden-Württemberg
Ravensburg, Campus Friedrichshafen
Academic Program Director
Automotive Electronics and
Mechatronic Systems

Editorial staff
Dipl.-Ing. Karl-Heinz Dietsche

Type setting
Dipl.-Ing. Karl-Heinz Dietsche

Translation
STAR Deutschland GmbH
Member of STAR Group

11th Edition, revised and extended,
January 2022

All rights reserved

Printed in Germany
Imprimé en Allemagne

Distribution
John Wiley & Sons Ltd
The Atrium
Southern Gate, Chichester
West Sussex PO19 8SQ, England
Phone (+44) 1243 779777
Internet: www.wiley.com
Email: customer@wiley.co.uk

ISBN 978-1-119-91190-6

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Authors
Dipl.-Ing. Karl-Heinz Dietsche
Prof. Dr.-Ing. Konrad Reif and
approx. 280 authors from industry
and the university and college sector

Technical graphics
Bauer & Partner, Gesellschaft für
technische Grafik mbH, Stuttgart
Schwarz Technische Grafik, Leonberg

Reproduction, duplication and translation of this publication, including excerpts therefrom, is only to ensue with our prior written consent and with particulars of source. Figures, descriptions, schematic diagrams, and other data are for explanatory purposes and illustration of the text only. They cannot be used as the basis for the design, installation, or specification of products.

The comments reflect solely the personal views of the respective authors. Comments relating to legal standards make no claims of being complete and correct. The contents of the Handbook do not make any statement as to whether vehicles and vehicle component parts are approvable in an individual case and whether they comply with the applicable legal requirements in each case.
Subject to alteration and amendment.

The brand names given in the contents are only examples and do not represent the classification or preference for a particular manufacturer. Trademarks are not identified as such.

The following companies kindly placed text and picture matter and diagrams at our disposal:

Automotive Lighting Reutlingen GmbH

BASF, Ludwigshafen

Bosch Engineering GmbH, Abstatt

Brose Schließsysteme GmbH & Co. KG, Wuppertal

BMW Group, München

Daimler Truck AG, Stuttgart

Dr. Ing. h.c. F. Porsche AG, Weissach

Eberspächer Climate Control Systems GmbH & Co. KG, Esslingen

ETAS GmbH, Stuttgart

Gates GmbH, Aachen

IAV GmbH, Berlin

iwis motorsysteme GmbH & Co. KG, München

J. Eberspächer GmbH & Co. KG, Esslingen

Knorr-Bremse Systeme für Nutzfahrzeuge GmbH, Schwieberdingen

MAHLE Behr GmbH & Co. KG, Stuttgart

MAN Nutzfahrzeuge Gruppe

MANN+HUMMEL GmbH, Ludwigsburg

Mercedes-Benz AG, Stuttgart

Mercedes-Benz AG, Sindelfingen

Michelin Reifenwerke AG & Co. KGaA, Karlsruhe

Robert Bosch Automotive Steering GmbH, Schwäbisch Gmünd

Saint-Gobain Sekurit Deutschland GmbH & Co. KG, Herzogenrath

SEG Automotive Germany GmbH, Stuttgart

Trelleborg Sealing Solutions Germany GmbH, Stuttgart

Volkswagen AG, Wolfsburg

ZF Friedrichshafen AG, Friedrichshafen

Duale Hochschule Baden-Württemberg, Ravensburg, Campus Friedrichshafen

Friedrich-Alexander-Universität, Erlangen-Nürnberg

Hochschule Esslingen

Hochschule Karlsruhe – Technik und Wirtschaft

Hochschule München

Karlsruher Institut für Technologie (KIT)

Reinhold-Würth-Hochschule, Künzelsau

Rheinisch-Westfälische Technische Hochschule (RWTH), Aachen

Foreword to the 11th Edition

Automotive engineering has become an extremely complex field over the last few years and decades. It is becoming increasingly more difficult to command an overview of the entire field and to maintain constant access to the subjects which are significant to automotive engineering. Many of these new subjects have in the meantime been covered in great detail in the wealth of available specialist literature. However, for those readers who wish to approach one of these new subjects for the first time, the available literature is neither easily manageable nor is it readable within the available timeframe.

This is where the Automotive Handbook comes in useful. It is structured in such a way as to be easily accessible even to those readers who are new to any individual subject. The most important subjects relevant to automotive engineering have been compiled in a compact, easily understandable, and practically relevant form. This is possible because the contents have been written by experts who work at Bosch, at other automotive companies and suppliers, and in the university sector in the very specialist fields presented.

In this the 11th Edition in English many subjects have been included for the first time, completely revised or enhanced. The subjects revised have centered on electrification of the drive. The Handbook has however also been revised, expanded and updated in many other places to make the contents even more understandable to the reader. Thus, the subject of artificial intelligence, which is used in many fields, has been included. More than 260 pages have been added to this latest edition. In spite of the large number of contributing authors, every effort has been made to provide a uniform presentation and to maintain consistent classifications and nomenclature within its pages.

This individual chapters of this publication have been laid out gradually on receipt of the authors' manuscripts. Care has been taken to ensure that the graphics where possible are positioned in such a way that they appear with the accompanying text on the same page or the facing page and consequently both are visible to the reader in a single glance. Shifting the pagination at a later stage by one page would undo this text/graphic reference. This does however mean that there are blank pages in some places.

This latest edition could not have been completed without the outstanding support of many individuals. Firstly, we would like to thank the authors of the individual articles, who, with great care and patience, have succeeded in delivering on schedule chapters of the highest substance and quality. Finally, we would like to thank all those readers who have provided useful suggestions and advice on corrections.

Friedrichshafen and Karlsruhe, January 2022,

Scientific advisory board
and editorship

Contents

Motor vehicles
Overall motor-vehicle system 28
Definition of motor vehicle 28
Functional units of a motor vehicle 28
Classification of motor vehicles 29

Basic principles
Quantities and units 32
SI units 32
Legal units 35
Further units 41
Natural constants 44
Mathematical signs and symbols 45
Greek alphabet 45
Basic principles of mechanics 46
Rectilinear and rotary motion 46
Statics 49
Calculation of strength 51
Friction 53
Fluid mechanics 56
Hydrostatics 56
Flow mechanics 57
Vibrations and oscillations 60
Terms 60
Equations 62
Vibration reduction 63
Modal analysis 64
Acoustics 66
General terminology 66
Measured quantities for noise emissions 69
Measured quantities for noise immissions 70
Perceived sound levels 71
Optical technology 74
Electromagnetic radiation 74
Geometric optics 74
Wave optics 77
Lighting quantities 79
Laser technology 82
Optical fibers/waveguides 83
Thermodynamics 86
Basic principles 86
Laws of thermodynamics 88
Equations of state 89
Changes of state 91
Cycles and technical applications 91
Heat transfer 96
Electrical engineering 102
Basic principles of electricity 103
Electrical quantities in an electric circuit 104
Conductors, nonconductors, semiconductors 109
Laws 110
Electric power and work 112
Time-dependent current 114
Electric field 117
Magnetic field 118
Electromagnetic fields 125
Power in the alternating-current circuit 126
Wave propagation 128
Electric effects in metallic conductors 131
Electronics 138
Basic principles of semiconductor technology 138
Discrete semiconductor devices 141
Monolithic integrated circuits 157
Basic principles of electrical machines 158
Energy conversion 158
Systematics of rotating electrical machines 159
Definition of geometric quantities 159
Direct-current machines 160
Asynchronous machine 166
Synchronous machine 168
Switched reluctance motor 173
Electronic motors 176
Energy-efficiency classes 177
Three-phase current system 178
Chemistry 182
Elements 182
Chemical bonds 188
Substances 192
Substance concentrations 194
Reaction of substances 195
Electrochemistry 202

Mathematics and methods
Mathematics 204
Numbers 204
Functions 204
Equations in the plane triangle 210
Complex numbers 210
Coordinate systems 211
Vectors 213
Differential and integral calculus 215
Linear differential equations 218
Laplace transform 219
Fourier transform 220
Matrix calculus 222

Technical statistics 226
Descriptive statistics 226
Probability calculus 228
Characteristic value in quality assurance and in quality management 232
Finite-element method 234
Application 234
FEM examples 237
Control engineering 242
Terms and definitions 242
Control-engineering transfer elements 243
Designing a control task 244
Adaptive controllers 246
Artificial intelligence 248
Introduction to AI 248
Machine learning 250
Representation of information 253
Planning 254
IT security of AI systems 256

Materials
Materials 258
Material parameters 258
Material groups 270
Metallic materials 271
EN metallurgy standards 286
Magnetic materials 290
Nonmetallic inorganic materials 303
Composite materials 303
Plastics 306
Heat treatment of metallic materials 328
Hardness 328
Heat-treatment processes 331
Thermochemical treatment 337
Corrosion and corrosion protection 340
Corrosion processes 340
Types of corrosion 341
Corrosion testing 342
Corrosion protection 344
Deposits and coatings 348
Deposits 348
Coatings 353

Machine parts
Tolerances 354
Form deviations 354
Geometrical Product Specifications 354
Axles and shafts 359
Function and application 359
Dimensioning 359
Design 361
Springs 362
Basic principles 362
Metal springs 364
Friction bearings 370
Features 370
Hydrostatic friction bearings 370
Hydrodynamic friction bearings 371
Self lubricating bearings made of metal 374
Self lubricating bearings made of plastic 376
Rolling bearings 378
Application 378
General principles 378
Selection of rolling bearings 379
Calculation of load capacity 382
Seals 384
Seal technology 384
O-ring 386
Molded seal 393
Flat seals 395
Sealing caps 396
Radial shaft seal 398
Pneumatic seals 406
Two-component seals 408
Rod and piston sealing systems 411
Elastomer materials 415

Joining and bonding techniques
Detachable connections 426
Positive or form-closed joints 426
Frictional joints 430
Threaded fasteners 437
Snap-on connections on plastic components 447
Permanent joints 450
Welding 450
Soldering 454
Adhesive technologies 455
Riveting 457
Penetration-clinching processes 459

Vehicle physics
Vehicle motion 462
Coordinate system 462
Degrees of freedom of vehicle motion 462
Running resistances 464
Total running resistance 464
Individual running resistances 464
Motor-vehicle dynamics 470
Characteristics of vehicle dynamics 470
Dynamics of linear motion 471
Dynamics of lateral motion 479
Dynamics of vertical motion 486
Special operating dynamics for commercial vehicles 493
Driving-dynamics test procedures 498
Assessment of vehicle handling 498
Driving-dynamics test procedures as per ISO 499
Energy demand for drives 506
Fuel consumption of internal-combustion engines 506
Fuel consumption of hybrid vehicles 509
Range and consumption of electric vehicles 510
Well-to-Wheel analysis 512
Vehicle aerodynamics 516
Aerodynamic forces 516
Tasks of vehicle aerodynamics 519
Vehicle wind tunnels 524
Vehicle acoustics 530
Legal requirements 530
Development work on vehicle acoustics 532
Sources of noise 533
Sound design 535

Operating fluids
Lubricants 538
Terms and definitions 538
Engine oils 544
Transmission oils 548
Lubricating oils 549
Lubricating greases 549
Coolants for internal-combustion engines 553
Requirements 553
Composition of coolants 553
Types of radiator protection agents 554
Brake fluids 556
Requirements 556
Characteristics 557
Chemical composition 558
Fuels 560
Overview 560
Characteristics 562
Gasoline (petrol) 563
Diesel fuel 569
Methane 579
Liquefied petroleum gas 581
Hydrogen 582
Electricity-generated regenerative fuels 583
Ethers 584
Urea/water solution 590
Refrigerants for air-conditioning systems 591
Development up to today 591
Alternative possibilities 592

Drivetrain
Drivetrain with internal-combustion engine 594
Drivetrain elements 594
Functions of the drivetrain 594
Tractive-force curves 595
Drive types 598
Power take-up elements and vibration isolation 600
Single clutch 600
Dual clutch 602
Hydrodynamic torque converter 603
Vibration isolation 605
Step-up gear units 608
Components in manually shifted transmissions 608
Manually shifted transmissions 614
Automated manual transmissions 617
Dual-clutch transmissions 618
Automatic planetary transmissions 620
Continuously variable transmissions 622
Commercial-vehicle transmissions 624
Differentials 626
Final-drive differentials 626
Electronic transmission control 630
Functions and requirements 630
Transmission-control components 631
Shift-point control 632
Hydraulic control 634
Electrohydraulic actuators 635

Internal-combustion engines
Internal-combustion engines 642
Thermal engines 642
Mixture formation, combustion, emissions 646
Gasoline engine 646
Diesel engine 657
Mixed forms and alternative operating strategies 662
Charge cycle and supercharging 664
Charge cycle 664
Variable valve timing 667
Supercharging processes 671
Exhaust-gas recirculation 674
Reciprocating-piston engine 676
Components 676
Reciprocating-piston engine types 690
Crankshaft-assembly design 691
Tribology and friction 699
Empirical values and data for calculation 703
Exhaust-gas emissions 716
Historic measures 716
Constituents of engine exhaust gas 716
Basic principles of catalytic exhaust-gas treatment 718
Catalytic processes for exhaust-gas treatment 719
Belt drives 720
Friction belt drives 720
Positive belt drives 725
Chain drives 730
Overview 730
Timing chains 730
Sprockets 734
Chain-tensioning and chain-guide elements 734
Engine cooling 736
Air cooling 736
Water cooling 737
Intercooling (charge-air cooling) 743
Exhaust-gas cooling 744
Oil and fuel cooling 745
Modularization 747
Intelligent thermal management 749
Engine lubrication 750
Forced-feed lubrication system 750
Components 751
Thermal management 753
Air-intake systems and intake manifolds 754
Overview 754
Passenger-car air-intake system 755
Passenger-car intake manifolds 763
Motorcycle intake manifolds 768
Commercial-vehicle air-intake system 768
Commercial-vehicle intake manifolds 771
Turbochargers and superchargers 772
Superchargers (mechanically driven) 772
Exhaust-gas turbochargers 775
Complex supercharging systems 784
Exhaust-gas system 790
Purpose and design 790
Exhaust manifold 792
Catalytic converter 792
Particulate filter 793
Mufflers 794
Connecting elements 796
Acoustic tuning devices 797
Commercial-vehicle exhaust-gas systems 798

Management for spark-ignition engines
Management for spark-ignition engines 800
Description of the engine management system 800
System overview 801
Motronic versions for passenger cars 805
Motronic versions for motorcycles 809
Cylinder charge 812
Component parts 812
Control of cylinder charge 813
Recording of air charge 815
Cylinder shutoff 817
Fuel supply 820
Fuel supply and delivery with intake-manifold injection 820
Fuel supply and delivery with gasoline direct injection 822
Evaporative-emissions control system 824
Gasoline filter 825
Electric fuel pump 827
High-pressure pumps for gasoline direct injection 830
Fuel rail 833
Fuel-pressure regulator 835
Fuel-pressure attenuator 836
Mixture formation 838
Basic principles 838
Mixture-formation systems 840

Intake-manifold injection 841
Gasoline direct injection 843
Fuel injectors 846
Ignition 854
Basic principles 854
Moment of ignition 855
Ignition systems 860
Ignition coil 862
Spark plug 865
Catalytic exhaust-gas treatment 874
Catalytic converter 874
λ control 879
Particulate filter for gasoline engines 881
Motor sport 882
Requirements 882
Components 882
ECUs 884
Pure motor-sport developments 884

Gasoline-engine operation with alternative fuels
LPG operation 886
Applications 886
Design 887
LPG systems 889
Components 891
Engines fueled by natural gas 894
Applications 894
Design 896
Components 897
Engines fueled by alcohol 900
Applications 900
Ethanol as an alternative fuel 901
Flex-fuel concepts for the markets 903
Flex-fuel components 904

Management for diesel engines
Electronic Diesel Control 906
Description of the engine-management system 906
Requirements 907
Operating principle and architecture 908
Data processing 910
Application 915
Low-pressure fuel supply 916
Fuel supply and delivery 916
Diesel filter 922
Common-rail fuel-injection system 924
System overview 924
Injectors 929
High-pressure pumps 937
Rail 943
Time-controlled single-cylinder pump systems 944
Unit-injector system for passenger cars 944
Unit-injector system for commercial vehicles 946
Unit-pump system for commercial vehicles 947
Electronic control 947
Diesel distributor injection pumps 948
Axial-piston distributor pumps 948
Radial-piston distributor pumps 951
Fuel-injection system 953
Start-assist systems for diesel engines 954
Preheating systems for passenger cars and light commercial vehicles 954
Flame start systems for commercial-vehicle diesel engines 958
Grid heater for commercial-vehicle diesel engines 961
Exhaust-gas treatment 962
Catalytic oxidation 962
Filtration of particulates 963
Catalytic reduction of nitrous fumes 965
On-Board Diagnostics (OBD) 969

Electrification of the drive
Electrification of the drive 970
Features 970
Components of the electrified drive 972
Electrified ancillary systems 975
Regenerative braking system 976
Thermal management for electric vehicles 980
Charging infrastructure for electrified vehicles 982
Control of electric drives 984
Drivetrain of an electric vehicle 984
Vehicle control unit for electric vehicles 984
Further applications of the VCU for electric vehicles 989
Hybrid drives 990
Motivation 990
Functionalities of hybrid drives 992
Classification of hybrid drives 994
Classification based on drive structure 998
Incorporating the transmission in the hybrid drive 1007

Control of a hybrid drive 1010
Hybrid vehicle 1010
Vehicle control unit for hybrid vehicles 1010
Drivetrain operating strategy for hybrid vehicles 1012
Electric final-drive unit 1020
Functional component integration 1020
Vehicle integration and interfaces 1024
Performance classes and performance characteristic 1026
Adaptation for application 1028
Electric drives with fuel cell 1030
Features 1030
Design and operating principle of the fuel cell 1031
Functioning principle of the fuel-cell system 1034
Fuel cell in the drivetrain 1038
Electromobility for two-wheelers 1040
Features 1040
System architecture 1041

Emission-control and diagnosis legislation
Emission-control legislation 1042
Overview 1042
Emission-control legislation for passenger cars and light commercial vehicles 1043
Test cycles for passenger cars and light commercial vehicles 1057
Emission-control legislation for heavy commercial vehicles 1062
Test cycles for heavy commercial vehicles 1067
Emission legislation for motorcycles 1071
Exhaust-gas measuring techniques 1078
Exhaust-gas testing 1078
Exhaust-gas measuring devices 1081
Diesel smoke-emission test 1085
Evaporative-emission test 1087
Diagnostics 1090
Monitoring in driving mode 1090
On-board diagnostics 1092
OBD functions 1098
OBD requirements for heavy commercial vehicles 1101
On-board diagnostics for motorcycles 1105
ECU diagnostics and Service Information System 1106

Chassis systems
Chassis systems 1108
Overview 1108
Basic principles 1110
Characteristics 1113
Suspension 1124
Basic principles 1124
Types of spring 1127
Suspension systems 1132
Shock absorbers and vibration absorbers 1136
Shock absorbers 1136
Vibration absorbers 1144
Wheel suspensions 1146
Basic principles 1146
Kinematics and elastokinematics 1146
Basic categories of wheel suspensions 1148
Wheels 1154
Function and requirements 1154
Structure 1154
Design criteria 1159
Designation for passenger-car wheels 1159
Materials for wheels 1160
Manufacturing processes 1162
Wheel design variations 1165
Stress and testing of wheels 1167
Tires 1170
Function and requirements 1170
Tire construction 1171
Tire inflation pressure 1173
Tire tread 1174
Force transmission 1176
Tire grip 1177
Rolling resistance 1180
Tire designation 1182
EU tire label 1185
Winter tires 1186
Development of a tire 1187
Tire tests 1187
Tire-pressure monitoring systems 1188
Rotary seal for tire-pressure control 1190
Steering 1192
Purpose of automotive steering systems 1192
Steering-system requirements 1193
Types of steering box 1195
Power-assisted steering systems for passenger cars 1196

Power-assisted steering systems for commercial vehicles 1203
Brake systems 1206
Definitions and principles 1206
Legal regulations 1211
Structure and organization of brake systems 1214
Brake systems for passenger cars and light commercial vehicles 1216
Subdivision of passenger-car brake systems 1216
Components of the passenger-car brake system 1217
Electrohydraulic brake 1224
Integrated power brake 1226
Brake systems for commercial vehicles 1230
System overview 1230
Components of commercial-vehicle brake systems 1233
Electronically controlled brake system 1242
Continuous-operation brake systems 1246
Wheel brakes 1250
Disk brakes for passenger cars 1250
Disk brakes for commercial vehicles 1255
Drum brakes 1256

Chassis control
Wheel-slip control systems 1260
Function and requirements 1260
Control systems 1261
ABS/TCS systems for passenger cars 1265
ABS/TCS systems for commercial vehicles 1269
Driving-dynamics control system 1272
Function 1272
Requirements 1273
Operating principle 1274
Typical driving maneuver 1274
Structure of the overall system 1275
System components 1284
Special driving-dynamics control system for commercial vehicles 1287
Automatic brake-system operations 1292
Motorcycle Stability Control 1296
Application 1296
Riding physics of a two-wheeler 1296
Motorcycle ABS 1298
Motorcycle Traction Control 1302
Integrated driving-dynamics control systems 1304
Overview 1304
Functions 1304
System architecture 1307

Vehicle bodies
Vehicle bodies, passenger cars 1310
Introduction 1310
Milestones in automobile history 1310
Definitions 1312
Vehicle body and vehicle segments 1313
Vehicle systems 1315
Interface management in complete-vehicle development 1316
Process of vehicle-concept development 1317
Sectional elaboration of the vehicle concept 1321
Current trends of vehicle-concept development 1322
Vehicle bodies, commercial vehicles 1324
Classification of commercial vehicles 1324
Light utility vans 1325
Medium- and heavy-duty trucks and tractor units 1325
Buses 1330
Active and passive safety in commercial vehicles 1332
Lighting equipment 1334
Functions 1334
Regulations and equipment 1335
Light sources 1336
Motor-vehicle bulbs 1337
Main-light functions 1337
Main-light functions for Europe 1341
Main-light functions for the USA 1343
Definitions and terms 1345
Technical design variations of headlamps 1346
Additional headlamps 1353
Light functions 1355
Installation and regulations for signal lamps 1359
Technical design variations for lamps 1363
Headlamp leveling control 1367
Headlamp adjustment 1368
Automotive glazing 1374
The material properties of glass 1374
Automotive windshield and window glass 1375

Functional design glazing 1376
Windshield, rear-window and headlamp cleaning systems 1380
Front wiper systems 1380
Rear wiper systems 1388
Windshield and rear-window washer systems 1389
Headlamp cleaning systems 1390

Comfort and convenience
Passenger-compartment climate control 1392
Climate-control requirements 1392
Design and operating principle of the A/C unit 1392
Climate-control systems 1395
Climate control for hybrid and electric vehicles 1396
Cabin filters 1397
Auxiliary heater systems 1398
Comfort and convenience systems in the door and roof areas 1404
Power-window systems 1404
Sunroof systems 1406
Comfort and convenience functions in the passenger compartment 1408
Electrical seat adjustment 1408
Electrical steering-wheel adjustment 1409

Safety systems in the motor vehicle
Safety systems in the motor vehicle 1410
Phases of vehicle safety 1410
Active safety 1412
Passive safety 1413
Legal regulations and consumer tests 1414
Occupant-protection systems 1416
Functions 1416
Retention systems and actuators 1418
Occupant sensing 1422
Integral-safety systems 1424
Overview 1424
Solution approaches 1425
Outlook 1429

Vehicle security systems
Locking systems 1430
Task and design 1430
Access authorization 1431
Lock design 1432
Requirements 1434
Lock layout 1434
Safety functions 1438
Development history 1439
Vehicle access by means of digital key 1441
Anti-theft systems 1444
Electronic vehicle immobilizer 1444
Theft-alarm system 1448

Automotive electrics
12 V vehicle electrical systems 1450
Function and requirements 1450
Power of the consumers in the 12 V vehicle electrical system 1451
Design and operating principle 1452
Interaction of the components 1455
Electrical-system structures 1458
Electrical-system parameters 1460
Electrical energy management 1462
Vehicle electrical systems for commercial vehicles 1466
Requirements of truck electrical systems 1466
Schematic design of truck electrical systems 1467
Electrical-system concepts for highly and fully automated driving functions 1471
12 V starter batteries 1476
Requirements 1476
Battery design 1478
Charging and discharging 1480
Battery characteristics 1482
Battery types 1483
Using the battery 1487
Alternators 1490
Electric power generation 1490
Alternator design 1492
Voltage regulation 1496
Characteristic alternator values 1498
Operating conditions 1499
Efficiency 1501
Alternator versions 1502
Starters for passenger cars and light commercial vehicles 1506
Application 1506
Requirements 1506
Starter design and operation 1507
Configuration of a starter 1511

Triggering the starter 1512
Starters for heavy commercial vehicles 1514
Application 1514
Power requirements 1514
Operating principle 1515
Special requirements for commercial-vehicle starters 1517
Wiring harnesses and plug-in connections 1518
Wiring harnesses 1518
Plug-in connections 1520
Electromagnetic compatibility 1524
Requirements 1524
Interference emission and interference immunity 1525
EMC-oriented development 1528
EMC measuring techniques 1529
EMC test methods 1530
EMC simulation 1533
Legal requirements and standards 1534
Symbols and circuit diagrams 1538
Circuit symbols 1538
Circuit diagrams 1545
Terminal designations 1557

Electrics for electric and hybrid drives
Vehicle electrical systems for hybrid and electric vehicles 1560
Vehicle electrical systems for mild- and full-hybrid vehicles 1560
Vehicle electrical systems for externally rechargeable hybrid and electric vehicles 1563
Charging strategy 1564
Electrical machines for motor-vehicle propulsion 1566
Requirements 1566
Design 1569
Types of electrical machines 1571
Influencing factors on operational performance 1574
Losses in electrical machines 1579
Cooling of the electrical machine 1580
Inverters for electric drives 1582
Area of application 1582
Function and circuitry 1583
Activation and control 1586
Software and functional safety 1589
Equipment mechanics 1591
DC-DC converters 1594
Application 1594
Non-isolating DC-DC converters 1595
Isolating DC-DC converters 1598
On-board charger 1600
Application 1600
Requirements 1601
Charging power 1602
Basic design and operating principle 1603
Batteries for electric and hybrid drives 1606
Requirements 1606
Storage technologies 1607
Basic structure of a battery system 1610
Components of a lithium-ion battery system 1610
Charging method for lithium-ion batteries 1615
Recycling of lithium-ion vehicle batteries 1615
Supercapacitors 1616
Application 1616
Electric double-layer capacitors 1617
Hybrid supercapacitors 1618

Automotive electronics
Electronic control unit (ECU) 1620
Functions 1620
Requirements 1621
ECU components 1624
Data processing 1632
Assembly and connection technology 1636
Software 1638
Application 1638
Functional safety 1639
EOL programming 1639
Mechatronics 1640
Mechatronic systems and components 1640
Development methodology 1642
Outlook 1645
Automotive software engineering 1646
Motivation 1646
Design of software in motor vehicles 1647
Important standards for software in motor vehicles 1647
The development process 1650
Quality assurance in software development 1654
Workflows of software development in motor vehicles 1654
Modeling and simulation of software functions 1656

Design and implementation of software functions 1659
Integration and testing of software and ECUs 1660
Calibration of software functions 1662
Outlook 1664
Architecture of electronic systems 1666
General 1666
Development of E/E architectures 1674
Summary and outlook 1680
Vehicle communication networks 1682
Bus systems 1682
Technical principles 1684
Buses in motor vehicles 1692
CAN 1692
FlexRay 1697
LIN 1700
Ethernet 1703
PSI5 1705
MOST 1707
Serial Wire Ring 1711
Automotive sensors 1716
Automotive applications 1716
Basic principles of classic sensors 1718
Position and angular-position sensors 1724
rpm sensors 1738
Oscillation gyrometers 1743
Flow meters 1746
Acceleration and vibration sensors 1750
Pressure sensors 1756
Temperature sensors 1760
Torque sensor 1764
Force sensor 1765
Gas and concentration sensors 1766
Optoelectronic sensors 1776

Infotainment
Infotainment and cockpit solutions 1780
History 1780
Infotainment systems 1781
Smartphone connection to the infotainment system 1782
Display and control 1784
Interaction channels 1784
Instrumentation 1785
Display types 1789
Voice control 1793
Radio and TV reception in motor vehicles 1794
Wireless signal transmission 1794
Radio tuners 1797
Traffic telematics 1801
Transmission paths 1801
Standardization 1801
Information recording 1802

Driver assistance and sensors
Driver-assistance systems 1804
Introduction – driver assistance 1804
Ultrasonic sensor system 1812
Ultrasonic sensor 1812
Distance measurement with ultrasound 1813
Radar-sensor system 1816
Application 1816
Radar principle 1817
Variants of radar sensors 1822
Homologation 1822
Lidar-sensor system 1824
Functions and applications 1824
Operating principle 1824
Properties and use 1827
Measurement principles 1828
Methods for increasing the number of pixels 1834
Video sensors 1838
Application 1838
Image sensor 1839
Objective 1842
Computer vision 1844
Sensor-data fusion 1860
Introduction 1860
Sensors 1861
Fusion types 1862
AI-based sensor-data fusion 1865

Driver-assistance systems
Vehicle navigation 1872
Navigation systems 1872
Functions of navigation 1872
Digital map 1875
Night-vision systems 1876
Applications 1876
Far-infrared systems 1876
Near-infrared systems 1877
Parking and maneuvering systems 1882
Application 1882
Ultrasonic parking aid 1882
Ultrasonic parking assistant 1884
Video systems 1886
Adaptive Cruise Control 1890
Function 1890
Design and function 1890
Control algorithms 1892

Area of application and functional expansions 1893
Current ongoing developments 1895
Information and warning systems 1896
Rearview-camera system 1896
Multi-camera system 1897
Digital exterior-mirror system 1898
Road-sign recognition 1899
Driver-drowsiness detection 1900
Rear cross-traffic warning 1900
Turn collision warning for trucks 1901
Moving-off information system for heavy commercial vehicles 1902
Wrong-way driver warning 1904
Lane assistance 1906
Lane departure warning 1906
Lane-keeping support 1908
Lane centering assist for heavy trucks 1908
Emergency lane-keeping support for heavy trucks 1909
Roadworks assistant 1909
Bottleneck/constriction assistant 1910
Lane-change assistant 1912
Blind-spot assistant 1912
Lane-change assistant 1913
Evasive steering support 1914
Further developments 1914
Emergency-braking systems 1916
Emergency-braking systems in longitudinal traffic 1916
Maneuver emergency braking 1919
Automatic emergency braking for unprotected road users 1920
Intersection assistant 1924
Motivation 1924
Left-turn assistant 1925
Cross-traffic assistant 1926
Traffic-light and stop-sign assistants 1927
Intelligent headlamp control 1928
Motivation 1928
System designs 1928
Advanced rider-assistance systems for two-wheelers 1930
Adaptive cruise control 1930
Forward collison warning 1931
Blind-spot detection 1931
Interior-monitoring systems 1932
Camera-based interior monitoring 1932

Future of automated driving
Future of automated driving 1934
Automation levels 1934
Obstacles to automated driving 1935
Stages on the road to autonomous driving 1936

Appendices
Technical terms 1940
Abbreviations 2031

Authors

Unless otherwise stated,
the authors are employees of Bosch.

Motor vehicles
Overall motor-vehicle system
Dipl.-Ing. Karl-Heinz Dietsche

Basic principles
Quantities and units
Dipl.-Ing. Karl-Heinz Dietsche
Prof. Dr. rer. nat. Susanne Schandl,
Duale Hochschule Baden-Württemberg,
Ravensburg, Campus Friedrichshafen

Basic principles of mechanics
Prof. Dr.-Ing. Horst Haberhauer,
Hochschule Esslingen

Fluid mechanics
Prof. Dr.-Ing. Horst Haberhauer,
Hochschule Esslingen

Vibrations and oscillations
Dipl.-Ing. Sebastian Loos,
Rheinisch-Westfälische Technische
Hochschule (RWTH) Aachen

Acoustics
Dipl.-Ing. Hans-Martin Gerhard,
Dr. Ing. h.c. F. Porsche AG, Weissach

Optical technology
Dr. rer. nat. Stefanie Mayer

Thermodynamics
Dr.-Ing. Ingo Stotz

Electrical engineering
Dipl.-Ing. Herbert Bernstein,
Freelance author
Dr.-Ing. Hans Roßmanith,
Friedrich-Alexander-Universität
Erlangen-Nuremberg
Prof. Dr.-Ing. Klemens Gintner,
Hochschule Karlsruhe – Technik und
Wirtschaft

Electronics
Prof. Dr.-Ing. Klemens Gintner,
Hochschule Karlsruhe – Technik und
Wirtschaft
Dr. rer. nat. Ulrich Schaefer

Basic principles of electrical machines
Prof. Dr.-Ing. Jürgen Ulm,
Reinhold-Würth-Hochschule, Künzelsau
Reinhardt Erli, M. Sc.,
Reinhold-Würth-Hochschule, Künzelsau

Chemistry
Dr. rer. nat. Jörg Ullmann

Mathematics and methods
Mathematics, Technical statistics
Prof. Dr.-Ing. Matthias E. Rebhan,
Hochschule München

Finite-element method
Prof. Dipl.-Ing. Peter Groth,
Hochschule Esslingen

Control engineering
Dr.-Ing. Wolf-Dieter Gruhle,
ZF Friedrichshafen AG, Friedrichshafen

Artificial intelligence
Dr. phil. Christoph Peylo
Dr.-Ing. Bastian Bischoff
Dr.-Ing. Mathias Bürger

Materials
Materials
Dr.-Ing. Hagen Kuckert
Dr. rer. nat. Jörg Ullmann
Dr. rer. nat. Witold Pieper
Dr. rer. nat. Waldemar Draxler
Dipl.-Ing. Angelika Schubert
Dipl.-Ing. Gert Lindemann
Dr.-Ing. Carsten Tüchert
Dr.-Ing. Sven Robert Raisch
Dr.-Ing. Reiner Lützeler
Dr. rer. nat. Jörg Bettenhausen
Dr.-Ing. Gerrit Hülder
Dipl.-Ing. Cornelius Gaida
Dipl.-Phys. Klaus-Volker Schütt

Heat treatment of metallic materials
Dr.-Ing. Thomas Waldenmaier
Dr.-Ing. Jochen Schwarzer

Corrosion and corrosion protection
Dipl.-Ing. (FH) Thomas Jäger

Deposits and coatings
Dr. Helmut Schmidt
Dipl.-Ing. (FH) Hellmut Schmid
Dipl.-Ing. (FH) Susanne Lucas

Machine parts
Tolerances, axles and shafts, springs, friction bearings
Prof. Dr.-Ing. Horst Haberhauer,
Hochschule Esslingen

Roller bearings
Dr.-Ing. Zhenhuan Wu

Seals
Dipl.-Ing. (FH) Ulrich Schmid
Dipl.-Ing. Sascha Bürkle
Thorsten Nitze, M. Eng.
Claudius Göller, MBE (DHBW)
Dipl.-Ing. Wolfgang Böhler
Frank Kleemann
Sergio Amorim, M. Sc.
Felix Schädler
All authors of the chapter "Seals" from Trelleborg Sealing Solutions Germany GmbH, Stuttgart.

Joining and bonding techniques
Detachable connections
Prof. Dr.-Ing. Horst Haberhauer,
Hochschule Esslingen
Dipl.-Ing. Rolf Bald
Dr.-Ing. Sven Robert Raisch

Permanent joints
Dr.-Ing. Knud Nörenberg,
Volkswagen AG, Wolfsburg
Dr. rer. nat. Patrick Stihler

Vehicle physics
Vehicle motion, running resistances
Dipl.-Ing. Marc Birk,
Mercedes-Benz AG, Stuttgart

Motor-vehicle dynamics
Dipl.-Ing. Marc Birk,
Mercedes-Benz AG, Stuttgart
Prof. Dr. rer. nat. Ludger Dragon,
Mercedes-Benz AG, Sindelfingen
Dr. Michael Hilgers,
Daimler Truck AG
Dr. Darko Meljnikov,
Daimler Truck AG

Driving-dynamics test procedures
Dr. Michael Hilgers,
Daimler Truck AG
Dr. Darko Meljnikov,
Daimler Truck AG

Energy demand for drives
Dipl.-Ing. Marc Birk,
Mercedes-Benz AG, Stuttgart
Dipl.-Ing. Martin Rauscher

Vehicle aerodynamics
Dipl.-Ing. Michael Preiß,
Dr. Ing. h.c. F. Porsche AG, Weissach

Vehicle acoustics
Dipl.-Ing. Hans-Martin Gerhard,
Dr. Ing. h.c. F. Porsche AG, Weissach

Operating fluids
Lubricants
Dr. rer. nat. Gerd Dornhöfer

Coolants for internal-combustion engines
Dr. rer. nat. Oliver Mamber,
MAHLE Behr GmbH & Co. KG, Stuttgart

Brake fluids
Dr. rer. nat. Michael Hilden
Dr. rer. nat. Harald A. Dietl,
BASF Ludwigshafen

Fuels
Dr. rer. nat. Jörg Ullmann

Urea/water solution
Thilo Schulz, M. Sc.
Dipl.-Ing. (FH) Manfred Fritz

Refrigerants for air-conditioning systems
Prof. Dr.-Ing. Stephan Engelking,
Duale Hochschule Baden-Württemberg,
Ravensburg, Campus Friedrichshafen

Drivetrain
Drivetrain with internal-combustion engine
Dipl.-Ing. Jürgen Wafzig, M. Sc.,
ZF Friedrichshafen AG, Friedrichshafen
Dipl.-Ing. Karl-Heinz Dietsche
Dipl.-Ing. (BA) Rolf Lucius Dempel

Power take-up elements
Dipl.-Ing. Jürgen Wafzig, M. Sc.,
ZF Friedrichshafen AG, Friedrichshafen
Dipl.-Ing. Karl-Heinz Dietsche
Dipl.-Ing. (BA) Rolf Lucius Dempel

Step-up gear units
Dipl.-Ing. Jürgen Wafzig, M. Sc.,
ZF Friedrichshafen AG, Friedrichshafen
Dipl.-Ing. Karl-Heinz Dietsche
Dipl.-Ing. (BA) Rolf Lucius Dempel
Ard van de Wiel, B Eng,
Bosch Transmission Technology B.V

Differentials
Dipl.-Ing. Jürgen Wafzig, M. Sc.,
ZF Friedrichshafen AG, Friedrichshafen

Electronic transmission control
Alexander Fuchs, M. Sc.
Dr.-Ing. Heiko Michel
Dipl.-Ing. (FH) Thomas Müller

Internal-combustion engines
Internal-combustion engines, mixture formation,
charge cycle and supercharging,
reciprocating-piston engine
Prof. Dr. sc. techn. Thomas Koch,
Karlsruher Institut für Technologie (KIT)
Tobias Michler, M. Sc.,
Karlsruher Institut für Technologie (KIT)
Dipl.-Ing. Heijo Oelschlegel,
Mercedes-Benz AG, Stuttgart
Dr.-Ing. Otmar Scharrer,
MAHLE Behr GmbH & Co. KG, Stuttgart

Exhaust-gas emissions
Dr.rer.nat. Christoph Osemann
Dr.-Ing. Hartmut Lüders

Belt drives
Dipl.-Ing. (FH). Mario Backhaus,
Gates GmbH, Aachen

Chain drives
Dr.-Ing. Thomas Fink,
iwis motorsysteme GmbH & Co. KG,
Munich
Dr.-Ing. Peter Bauer,
iwis motorsysteme GmbH & Co. KG,
Munich

Engine cooling
Dipl.-Ing. (FH) Ralf-Holger Schink,
MAHLE Behr GmbH & Co. KG, Stuttgart
Dr.-Ing. Otmar Scharrer,
MAHLE Behr GmbH & Co. KG, Stuttgart

Engine lubrication
Dipl.-Ing. Markus Kolczyk,
MANN+HUMMEL GmbH, Ludwigsburg
Dr.-Ing. Harald Banzhaf,
MANN+HUMMEL GmbH, Ludwigsburg

Air-intake systems and intake manifolds
Dipl.-Ing. Andreas Weber,
MANN+HUMMEL GmbH, Ludwigsburg
Dipl.-Ing. Andreas Pelz,
MANN+HUMMEL GmbH, Ludwigsburg
Dipl.-Ing. Markus Kolczyk,
MANN+HUMMEL GmbH, Ludwigsburg
Dipl. Ing. (FH) Alexander Korn,
MANN+HUMMEL GmbH, Ludwigsburg
Dipl. Ing. (FH) Matthias Alex,
MANN+HUMMEL GmbH, Ludwigsburg
Dr.-Ing. Matthias Teschner,
MANN+HUMMEL GmbH, Ludwigsburg
Dipl. Ing. (FH) Mario Rieger,
MANN+HUMMEL GmbH, Ludwigsburg
Dipl. Ing. Christof Mangold,
MANN+HUMMEL GmbH, Ludwigsburg
Dipl.-Ing. Hedwig Schick,
MANN+HUMMEL GmbH, Ludwigsburg

Turbochargers and superchargers
Dr.-Ing. Gunter Winkler,
BMTS Technology GmbH & Co. KG,
Stuttgart

Exhaust-gas system
Dr. rer. nat. Rolf Jebasinski,
J. Eberspächer GmbH & Co. KG,
Esslingen

Management for gasoline engines
Management for gasoline engines
Dipl.-Ing. Armin Hassdenteufel
Dr.-Ing. Henning Heikes

Cylinder charge
Dr.-Ing. Martin Brandt
Dr.-Ing. Henning Heikes
Dipl.-Ing. (FH) Thomas Kortwittenborg
Dipl.-Ing. Frank Walter

Fuel supply
Dipl.-Ing. Timm Hollmann
Dipl.-Ing. Thomas Herges
Dipl.-Ing. Johannes Högl
Dipl.-Ing. Karsten Scholz
Dipl.-Ing. Michael Kuehn
Dipl.-Ing. Zdenek Liner
Dr.-Ing. Thomas Kaiser
Dipl.-Ing. Serdar Derikesen
Dipl.-Ing. Uwe Müller
Dipl.-Ing. (FH) Horst Kirschner

Mixture formation
Dipl.-Ing. Andreas Binder
Dipl.-Ing. Anja Melsheimer
Dipl.-Ing. Markus Gesk
Dipl.-Ing. Andreas Glaser
Dr.-Ing. Tilo Landenfeld

Ignition
Dr.-Ing. Martin Brandt
Dipl.-Ing. Walter Gollin
Dipl.-Ing. Werner Häming
Dipl.-Ing. Tim Skowronek
Dr.-Ing. Grit Vogt
Dr.-Ing. Matthias Budde

Catalytic exhaust-gas treatment
Dipl.-Ing. Klaus Winkler
Dipl.-Ing. Detlef Heinrich
Dr.-Ing. Frank Meier

Motor sport
Dipl.-Red. Ulrich Michelt,
Bosch Engineering GmbH, Abstatt

Gasoline-engine operation with alternative fuels
LPG operation
Dipl.-Ing. Iraklis Avramopoulos,
IAV GmbH, Berlin

Engines fueled by natural gas
Dipl.-Ing. (FH) Thorsten Allgeier

Engines fueled by alcohol
Dipl.-Ing. Andreas Posselt

Management for diesel engines
Management for diesel engines
Dipl.-Ing. Klaus Schwarze

Low-pressure fuel supply
Dipl.-Ing. (FH) Stefan Kieferle
Dr.-Ing. Thomas Kaiser
Dipl.-Ing. Serdar Derikesen

Common-rail injection system
Dipl.-Ing. Felix Landhäußer
Dipl.-Ing. (FH) Andreas Rettich
Dr. rer. nat. Matthias Schnell
Dipl.-Ing. Thilo Klam
Dr.-Ing. Holger Rapp
Anees Haider Bukhari, B. Eng, MBA
Dipl.-Ing. (FH) Herbert Strahberger

Time-controlled single-cylinder pump system
Dipl.-Ing. (BA) Jürgen Crepin,
ETAS GmbH, Stuttgart

Diesel-distributor injection pumps
Dipl.-Ing. (BA) Jürgen Crepin,
ETAS GmbH, Stuttgart

Start-assist systems
Dipl.-Ing. (FH) Steffen Peischl
Dipl.-Ing. Friedrich Schmid,
Mercedes-Benz AG, Stuttgart

Exhaust-gas treatment
Dr. rer. nat. Christoph Osemann
Dr.-Ing. Hartmut Lüders

Electrification of the drive
Electrification of the drive
Dipl.-Ing. Karl-Heinz Dietsche
Dr. rer. nat. Anselm Stüken
Benjamin Heinz, B. Eng.
Martin Kuehnemund, M. Sc.
Dr. Thomas Weil

Control of electric drives
Dipl.Ing. (FH) Harald Paulhart
Dipl.Ing. (FH) Arno Hafner

Hybrid drives
Dipl.-Ing. Thomas Huber
Dipl.-Ing. Rasmus Frei
Dipl.-Ing. Karl-Heinz Dietsche

Control of hybrid drives
Ph.D. James Girard
Dipl.-Ing. (FH) Arno Hafner
Dipl.-Ing. (FH) Harald Paulhart

Electric final drive
Dipl.-Ing (FH) Matthias Herrmann
Dr.-Ing. Andreas Wuensch
Dipl.-Ing (FH) Bernd Spielmann

Electric drives with fuel cells
Dipl.-Ing. (FH) Jan-Michael Grähn
Dr. rer. nat. Michael G. Marino
Tiffany Mittelstenscheid, M. Sc.

Electromobility for two-wheelers
Sven Digele, M.Sc.

Emission-control and diagnosis legislation
Emission-control Legislation
Dr.-Ing. Matthias Tappe
Dipl.-Ing. Michael Bender

Exhaust-gas measuring techniques
Dipl.-Phys. Martin-Andreas Drühe
Felix Reinke, B.Sc.
Dipl.-Ing. Andreas Weiß
Dipl.-Ing. (BA) Marc Rottner
Dipl.-Ing. Andreas Kreh
Dipl.-Ing. Bernd Hinner
Dr.-Ing. Matthias Tappe
Dr.-Ing. Christelle Oediger

Diagnostics
Dr.-Ing. Markus Willimowski
Dr. Richard Holberg
Dr. rer. nat. Hauke Wendt
Dipl.-Ing. Sani Dzeko

Chassis systems
Chassis systems
Univ.-Prof. Dr.-Ing. Ralph Mayer,
Technische Universität Chemnitz
Prof. Dr. rer. nat. Ludger Dragon,
Mercedes-Benz AG, Sindelfingen

Suspension, shock absorbers and vibration absorbers, wheel suspension
Univ.-Prof. Dr.-Ing. Ralph Mayer,
Technische Universität Chemnitz
Dipl.-Ing. Fridtjof Körner,
Technische Universität Chemnitz

Wheels
Dipl.-Ing. Martin Lauer,
Mercedes-Benz AG, Sindelfingen
Dipl.-Ing. Werner Hann,
Mercedes-Benz AG, Sindelfingen
Dipl.-Hdl. Martin Bauknecht,
MAN Nutzfahrzeuge Group

Tires
Dipl.-Ing. Dirk Vincken,
Agentur für Text&Bild, Eurasburg
Dipl.-Ing. Reimund Müller,
Michelin Reifenwerke AG & Co. KGaA
Karlsruhe
Thilo Balcke,
Michelin Reifenwerke AG & Co. KGaA
Karlsruhe
Dipl.-Ing. Norbert Polzin
Frank Kleemann,
Trelleborg Sealing Solutions

Steering
Dipl.-Ing. Peter Brenner,
Robert Bosch Automotive Steering GmbH
Schwäbisch Gmünd

Brake systems
Dr. rer. nat. Jürgen Bräuninger
Werner Schneider

Brake systems for passenger cars
Werner Schneider
Dipl.-Ing. Isabell Maier
Dipl.-Ing. Andreas Burg
Dipl.-Ing. Andreas Mayer
Dipl.-Ing. Hubertus Wienken
Dipl.-Ing. Frank Bährle-Miller
Dipl.-Ing. Bernhard Kant
Dipl.-Ing. Urs Bauer

Brake systems for commercial vehicles
Werner Schneider
Dr.-Ing. Falk Hecker,
Knorr-Bremse SfN, Schwieberdingen
Dipl.-Ing. Frank Schwab,
Knorr-Bremse SfN, Schwieberdingen
Dr.-Ing. Dirk Huhn,
ZF Friedrichshafen AG, Friedrichshafen

Wheel brakes
Werner Schneider
Dipl.-Ing. Andreas Burg
Dipl.-Ing. Andreas Mayer
Dr. Christof Gente
Dr.-Ing. Falk Hecker,
Knorr-Bremse SfN, Schwieberdingen

Chassis control
Wheel-slip control systems
Dr.-Ing. Falk Hecker,
Knorr-Bremse SfN, Schwieberdingen
Dipl.-Ing. Frank Schwab,
Knorr-Bremse SfN, Schwieberdingen

Driving-dynamics control system
Dr.-Ing. Gero Nenninger
Dipl.-Ing. (FH) Jochen Wagner
Dr.-Ing. Falk Hecker,
Knorr-Bremse SfN, Schwieberdingen

Motorcycle Stability Control
Dr. rer. nat. Markus Lemejda
Dr.-Ing. Fevzi Yildirim

Integrated driving-dynamics control systems
Dr.-Ing. Michael Knoop

Vehicle bodies
Vehicle bodies, passenger cars
Dr.-Ing. Jochen Bisinger,
Mercedes-Benz AG
Dr. rer. pol. Dipl.-Ing. Thomas M. Neff,
Mercedes-Benz AG
Dipl.-Ing. (FH) Hubert Kazmaier,
Mercedes-Benz AG
Dipl.-Ing. Matthijs Ravestein,
Mercedes-Benz AG

Vehicle bodies, commercial vehicles
Dipl.-Ing. Georg Stefan Hagemann,
Mercedes-Benz AG, Stuttgart

Lighting equipment
Dipl.-Ing. Doris Boebel,
Automotive Lighting Reutlingen GmbH
Dipl.-Ing. Gert Langhammer,
Automotive Lighting Reutlingen GmbH
Dr.-Ing. Michael Hamm,
Audi AG, Ingoldstadt

Automotive glazing
Caroline Hermsdorff, M.A.
Saint-Gobain Sekurit Deutschland
GmbH & Co.KG, Herzogenrath

Windshield, rear-window and headlamp cleaning systems
Dr. techn. Günter Kastinger

Comfort and convenience
Passenger-compartment climate control
Dipl.-Ing. Peter Kroner,
MAHLE Behr GmbH & Co. KG, Stuttgart
Dipl.-Ing. (FH) Thomas Feith,
MAHLE Behr GmbH & Co. KG, Stuttgart
Dipl.-Ing. Günter Eberspach,
Eberspächer Climate Control Systems
GmbH & Co. KG, Esslingen
Ulrich Karl Weber,
Eberspächer Climate Control Systems
GmbH & Co. KG, Esslingen
Dipl.-Betriebsw. Marcel Bonnet

Comfort and convenience systems in the door and roof areas
Dipl.-Ing. (FH) Walter Haußecker
Dipl.-Ing. (FH) Siegfried Reichmann

Comfort and convenience functions in the passenger compartment
Mattias Hallor, M. Sc.

Safety systems in the motor vehicle
Safety systems in the motor vehicle
Dipl.-Ing. Robert Krott
Dipl.-Ing Karl-Heinz Dietsche

Occupant-protection systems
Dipl.-Ing. Robert Krott

Integral-safety systems
Dr.-Ing. Dagobert Masur

Vehicle security systems
Locking systems
Dr.-Ing. Ulrich Nass,
Brose Schließsysteme GmbH & Co. KG,
Wuppertal
Michael Bergmann, M. Eng.
Emre Inal, M. Sc.

Anti-theft systems
Dipl.-Ing. (FH) Manuel Wurm,
ZF Friedrichshafen AG, Friedrichshafen
Dipl.-Ing. (FH) U. Götz

Automotive electrics
12 V vehicle electrical systems
Dipl.-Ing. Eberhard Schoch
Dipl.-Ing. Clemens Schmucker
Dipl.-Ing. Markus Beck

Vehicle electrical systems for commercial vehicles
Dipl.-Ing. Christian Bohne
Dr.rer.nat. Felix Hoos
Dipl.-Ing. Matthias Horn

12 V starter batteries
Dr. rer. nat. Richard Aumayer

Alternators
Dipl.-Ing. Reinhard Meyer,
SEG Automotive Germany GmbH

Starters for passenger cars and light commercial vehicles
Dipl.-Ing. Oliver Capek,
SEG Automotive Germany GmbH
Andreas Brosche, M. Sc.,
SEG Automotive Germany GmbH

Starters for heavy commercial vehicles
Dipl.-Ing. Henning Stoecklein,
SEG Automotive Germany GmbH

Wiring harnesses and plug-in connections
Dr.-Ing. Eckhardt Philipp

Electromagnetic compatibility (EMC)
Dr.-Ing. Wolfgang Pfaff

Symbols and circuit diagrams
Dipl.-Ing. Karl-Heinz Dietsche

Electrics for electric and hybrid drives
Vehicle electrical systems for hybrid and electric vehicles
Dr.-Ing. Hans-Peter Groeter

Electric machines as drives
Dr.-Ing. Marcus Alexander
Dr.-Ing. Arndt Kelleter
Dr.-Ing. Stephan Usbeck
Dipl.-Ing Uwe Knappenberger
Dr.-Ing. Marc Brück

Inverters for electric traction drives
Dr.-Ing. Hans-Peter Groeter

DC-DC converters
Dipl.-Ing. (FH) Wolfgang Haas

On-board charger
Dr.-Ing. Christoph van Booven

Batteries for electric and hybrid drives
Dr. rer. nat. Richard Aumayer

Supercapacitors
Dr. Volker Doege

Automotive electronics
ECUs
Dipl.-Ing. Klaus Schwarze
Dipl.-Ing. Axel Aue

Mechatronics
Dipl.-Ing. Hans-Martin Heinkel
Dr.-Ing. Klaus-Georg Bürger

Automotive software engineering
Dipl.-Ing. (BA) Jürgen Crepin,
ETAS GmbH, Stuttgart
Dr. rer. nat. Kai Pinnow,
ETAS GmbH, Stuttgart
Dipl.-Ing. Jörg Schäuffele,
ETAS GmbH, Stuttgart

Architecture of electronic systems
Dr.-Ing. Ralf Machauer,
Bosch Engineering GmbH, Abstatt
Andreas Hörtling, M. Sc.,
Bosch Engineering GmbH, Abstatt
Andreas Ehrhart, M. Sc.,
Bosch Engineering GmbH, Abstatt

Vehicle communication networks
Dr. rer. nat. Harald Weiler
Dr.-Ing. Tobias Lorenz
Dipl.-Ing. Karl-Heinz Dietsche

Buses in motor vehicles
Dr. rer. nat. Harald Weiler
Dr.-Ing. Tobias Lorenz
Dipl.-Ing. Oliver Prelle
Dipl.-Ing. Dieter Thoss

Automotive sensors
Dr.-Ing. Erich Zabler
Dipl.-Ökon. Frauke Ludmann
Dr. rer. nat. Peter Spoden
Dipl.-Ing. (FH) Cyrille Caillié
Dr.-Ing. Uwe Konzelmann
Dr.-Ing. Tilmann Schmidt-Sandte

Dr.-Ing. Reinhard Neul
Dr. Berndt Cramer
Dipl.-Ing. Dipl.-Wirt.-Ing. Nils Kaiser
Dipl.-Ing. Karl-Heinz Dietsche

Infotainment
Infotainment and cockpit solutions
Dipl.-Ing. Karl-Heinz Dietsche

Display and control
Prof. Dr.-Ing. Peter Knoll
Dipl.-Ing. Karl-Heinz Dietsche

Radio and TV reception in motor vehicles
Dr.-Ing. Jens Passoke

Traffic telematics
Dr.-Ing. Michael Weilkes

Driver assistance and sensor system
Driver-assistance systems
Dr.-Ing. Thomas Michalke
Dr.-Ing. Frank Niewels
Dipl.-Math. (FH) Thomas Lich
Dr.-Ing. Thomas Maurer

Ultrasonic sensor system
Prof. Dr.-Ing. Peter Knoll

Radar-sensor system
Dipl.-Ing. Joachim Selinger,
M. Sc. (University of Colorado)
Dr.-Ing. Michael Schoor

Lidar-sensor system
Dr.-Ing. Jan Sparbert

Video sensors
Dr. sc. Moritz Eßlinger
Christian Schwarz, M. Sc.
Dr.-Ing. Stjepan Dujmovic

Sensor-data fusion
PD Dr.-Ing. habil. Alexandru Paul
Condurache
Dr.-Ing. André Treptow

Driver-assistance systems
Vehicle navigation
Dipl.-Ing. Ernst Peter Neukirchner

Night-vision systems
Prof. Dr.-Ing. Peter Knoll

Parking and maneuvering systems
Prof. Dr.-Ing. Peter Knoll

Adaptive Cruise Control
Dipl.-Ing. Gernot Schröder
Prof. Dr.-Ing. Peter Knoll

Information and warning systems
Dipl.-Ing. Karl-Heinz Dietsche
Dr. Benjamin Schoen

Lane assistance
Dipl.-Ing. Karl-Heinz Dietsche
Dr.-Ing. Thomas Michalke
Dr. rer. nat. Lutz Bürkle

Lane-change assistant
Dipl.-Ing. Karl-Heinz Dietsche
Dr.-Ing. Thomas Michalke
Dr. rer. nat. Lutz Bürkle

Emergency-braking systems
Dipl.-Ing. Karl-Heinz Dietsche
Dr.-Ing. Thomas Gussner
Dr.-Ing. Steffen Knoop
Dr.-Ing. Falk Hecker,
Knorr-Bremse SfN, Schwieberdingen

Intersection assistant
Dipl.-Ing. Karl-Heinz Dietsche
Dr. rer. nat. Wolfgang Branz
Dr.-Ing. Rüdiger Jordan

Intelligent headlamp control
Dipl.-Ing. Doris Boebel,
Automotive Lighting Reutlingen GmbH
Dipl.-Ing. (FH) Bernd Dreier,
Automotive Lighting Reutlingen GmbH
Dipl.-Ing. Karl-Heinz Dietsche

Advanced rider-assistance systems for two-wheelers
Dipl.-Ing. Karl-Heinz Dietsche

Interior-monitoring systems
Dipl.-Ing. Karl-Heinz Dietsche

Future of automated driving
Future of automated driving
Holger Scharf

Overall motor-vehicle system

Definition of motor vehicle

The International Convention on Road Signs and Signals [1] defines a motor or power-driven vehicle as a self-propelled road vehicle. Rail-borne vehicles and country-specifically mopeds which are not treated as motorcycles are exceptions. A second definition in this convention limits the term motor vehicle to those vehicles which are used for carrying persons or goods or for drawing on the road vehicles used for the carriage or persons or goods. Electrically powered non-rail-borne trolley-buses are included in this definition. This definition however does not cover vehicles, such as agricultural tractors, which are only incidentally used for carrying persons or goods.

An appropriate definition is established for example in the German Road Traffic Act (§1 StVG [2]) and in the laws of other countries.

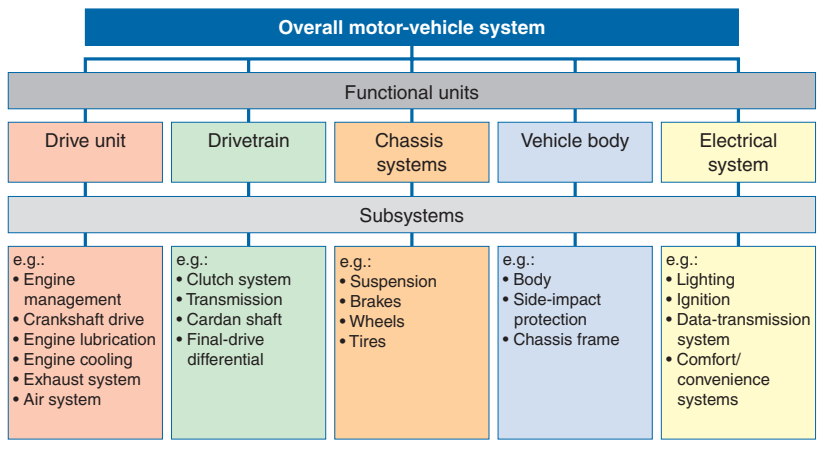
Functional units of a motor vehicle

A motor vehicle consists of a number of functional units (Figure 1). These are usually classified into drivetrain, drive unit, chassis, vehicle body, and electrical system. This classification is however not standardized.

The functional units are composed of the subsystems, which are interconnected. But there is also an interaction between the functional units. Thus, for example, the drive unit acts on the transmission unit with the transmission subsystem, which drives the wheels of the chassis functional unit via the drive shafts and the final-drive differential.

Electrical and electronic systems have gradually found their way into all areas of a motor vehicle. They have replaced mechanical systems or else created completely new possibilities. Electrification of the drive and further developments in the field of automated driving have major impacts on motor-vehicle development.

Figure 1: System network of a motor vehicle



Classification of motor vehicles

Motor vehicles include motor cars, but also single-track motorcycles (Figure 2). In the case of two- or multiple-track motor cars, a distinction is made between passenger cars and commercial vehicles (or goods vehicles). Commercial vehicles included trucks, motor buses, and tractors.

Determination of motor vehicles

Passenger cars

Passenger cars have at least four wheels. They are also called automobiles. Passenger cars serve to transport persons and additionally luggage, but also goods. A trailer can also be towed.

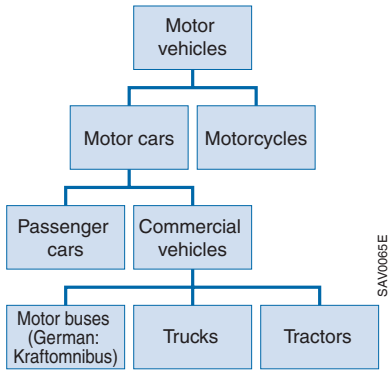
Commercial vehicles

Commercial vehicles serve to transport persons and goods and to tow trailer vehicles.

Motorcycles

Motorcycles are single-track motor vehicles, but may also have a sidecar attached.

Figure 2: Classification of motor vehicles



Country-specific classification

EC vehicle categories

Motor vehicles are classified according to their determination into vehicle categories. There are different classifications depending on the legislation. Table 1 provides an overview of the vehicle categories in Europe. This classification applies to the countries of the European Union and all other European countries.

Off-road vehicles are a special case. Vehicles in category N1 with a gross weight rating of not more than 2 tons and vehicles in category M1 are considered in accordance with Directive 87/403/EEC [3] to be off-road vehicles (off-road cars) if they are equipped as follows:

- They have at least one front axle and at least one rear axle, both of which can be driven simultaneously (all-wheel drive). The drive to one of the axles can be disengaged.
- A differential lock or a mechanism that has a similar effect must be fitted.
- As a solo vehicle (without a trailer) they must be able to climb a gradient of 30 % (verified by calculation).
- There are also further requirements for example with regard to ground clearance.

Table 1: EC vehicle categories

L	L vehicles	
	L1e-A	Powered cycle (pedelec)
	L1e-B	Two-wheel moped
	L2e	Three-wheel moped, sub-categorized by special use (-P, -U)
	L3e	Two-wheel motorcycle, sub-categorized by performance (-A1, -A2, -A3) and special use (-AxE, -AxT)
	L4e	Two-wheel motorcycle with sidecar (motorcycle combination), subcategories according to L3e
	L5e	Powered tricycle, sub-categorized by special use (-A, -B)
	L6e-A	Light on-road quad, sub-categorized by performance (-A, -B) and special use (-BU, -BP)
	L7e	Heavy on-road quad (max. effective power 15 kW, curb weight up to 400 kg, up to 600 kg for carriage of goods, in each case without batteries for electric vehicles), sub-categorized by special use

SUVs (sport utility vehicles) are based in terms of design on off-road vehicles, but do not satisfy all the above-mentioned requirements. Off-road capability varies from model to model. In fact, they are for the most part not used in off-road applications at all.

CARB vehicle categories
Table 2 provides an overview of the vehicle categories of the CARB legislation. This applies to the US state of California and some other states.

EPA vehicle categories
All other states are covered by the EPA legislation. The EPA vehicle categories are listed in Table 3.

Table 1: EC vehicle categories (continued)

M	Motor vehicles for carriage of passengers with at least four wheels	
	M1	Vehicles with no more than eight seats (in addition to driver's seat) Carriage of standing passengers is not possible
	M2	Vehicles with more than eight seats having a GVW not exceeding 5 tons Carriage of standing passengers in addition to the sitting passengers is possible
	M3	Vehicles with more than eight seats having a GVW exceeding 5 tons Carriage of standing passengers in addition to the sitting passengers is possible
N	Motor vehicles for carriage of goods with at least four wheels	
	N1	Vehicles with a GVW not exceeding 3.5 tons
	N2	Vehicles with a GVW not exceeding 12 tons
	N3	Vehicles with a GVW exceeding 12 tons
O	Trailers for carriage of goods or passengers, including semi-trailers	
	O1	Trailers not exceeding 750 kg (light trailers)
	O2	Trailers not exceeding 3.5 tons
	O3	Trailers not exceeding 10 tons
	O4	Trailers exceeding 10 tons

Vehicle categories of other countries
Japanese legislation follows the classification of CARB. China and India follow the EU classification.

Body types and segments of passenger cars
Various demands are placed on passenger cars, depending on their use. This gives rise in all markets to different segments with different body types. For the M categories the following different body types exist [4]:
– AA Sedan/saloon,
– AB Fastback sedan/saloon,
– AC Station wagon/estate car,
– AD Coupé,
– AE Convertible sedan/saloon,
– AF Multipurpose vehicle (e.g., fire truck, ambulance),
– AG Passenger-car pick-up.

The classification of vehicles into segments is fluid since there is no official definition. This can be attributed among other things to the fact that over time new vehicle models have always been brought onto the market. The following designations have become established:
– mini,
– subcompact,
– lower mid-size category (compact category with transversally mounted front engine, fastback with tailgate),
– mid-size category,
– upper mid-size category,
– luxury class,
– SUV,

Table 2: CARB vehicle categories

PC	Passenger car for carriage of up to 12 persons	
LDT	Light-duty truck for transport purposes	
	LDT1	Vehicles with maximum LVW (loaded vehicle weight) of 3,750 pounds
	LDT2	Vehicles with maximum GVWR (gross vehicle weight rating) of 8,500 pounds
MDV	Medium-duty vehicle, GVWR of 8,501 to 14,000 pounds	
HDV	Heavy-duty vehicle with GVPR > 8500 pounds, except PC	

– off-road vehicle,
– van (small vans and large vans).

There are also vehicles with dedicated body types:
– SA Recreational vehicles,
– SB Armored,
– SC Ambulances,
– SD Hearses,
– SG Miscellaneous,
– SH Wheelchair-accessible.

Table 3: EPA vehicle categories

LDV	Light-duty vehicle (passenger car) for carriage of up to 12 persons	
LLDT	Light light-duty truck with GVWR < 6,000 lbs (gross vehicle weight rating) LDT1 and LDT2 is combined in LLDT	
HLDT	Heavy light-duty truck with GVWR > 6,000 lbs LDT3 and LDT4 is combined in HLDT	
LDT	Light-duty truck, vehicle with GVWR < 8,500 lbs or curb weight < 6,000 lbs or frontal surface area less than 45 square feet; for transport of property or carriage of more than 12 persons or with properties for off-road operation	
	LDT1	LLDT with LVW < 3,750 lbs (loaded vehicle weight)
	LDT2	LLDT with LVW > 3,750 lbs
	LDT3	HLDT with ALVW < 5,750 lbs
	LDT4	HLDT with ALVW > 5,750 lbs
Full Size Pickup Truck	Light truck with passenger compartment and open freight box, towing capability > 5,000 lbs, payload > 1,700 lbs	
MDPV	Medium-duty passenger vehicle, Heavy-duty vehicle with GVWR <10,000 lbs for carriage of up to 12 persons	
HDV	Heavy-duty vehicle with GVWR > 8,500 lbs, or curb weight > 6,000 lbs, or frontal surface area greater than 45 square feet	

References
[1] Rechtsinformationssystem des Bundes (RIS) – Übereinkommen über Straßenverkehrszeichen; https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10011543.
[2] Straßenverkehrsgesetz (StVG) [German Road Traffic Act] § 1 Zulassung (Type Approval).
[3] Council Directive 87/403/EEC of 25 June 1987 supplementing Annex I to Directive 70/156/EEC on the approximation of the laws of the Member States relating to the type-approval of motor vehicles and their trailers.
[4] Kraftfahrt-Bundesamt (Federal Motor Transport Authority): Verzeichnis zur Systematisierung von Kraftfahrzeugen und ihren Anhängern. As at: March 2020. https://www.kba.de.

Quantities and units

In order to be able to express the values of physical quantities, a system of units is required which serves as the yardstick for every measurement. The value of the physical quantity is taken as the product of numerical value and unit. Such a system of units is the SI system, which was established in 1960 by the 11th General Conference on Weights and Measures (Conférence Générale des Poids et Mesures, CGPM). It has since then been adopted by over 50 countries.

In Germany the management of units is by law under the control of the Physikalisch-Technische Bundesanstalt (PTB, based in Braunschweig), Germany's national metrology institute. International responsibility lies with the International Bureau of Weights and Measures (BIPM, Bureau International des Poids et Mesures) in Sèvres near Paris.

Further units which are to this day commonly used (e.g., liter, metric ton, hour, degree Celsius) are in Germany likewise permitted by law and are mentioned here as such.

Units permitted in other countries (e.g., inch, ounce, degree Fahrenheit) or obsolete units are discussed in a separate section.

Table 1: Base SI units

Base quantity and symbol		Base SI unit	
		Name	Symbol
Length	l	meter	m
Mass	m	kilogram	kg
Time	t	second	s
Electric current	I	ampere	A
Thermodynamic temperature	T	kelvin	K
Amount of substance	n	mole	mol
Luminous intensity	I	candela	cd

SI units

SI means "Système International d'Unités" (International System of Units). The system is laid down in ISO 80000 [1] (ISO: International Organization for Standardization) and for Germany in DIN 1301 [2] (DIN: German Institute for Standardization).

Table 1 lists the seven base SI units.

Definition of the base SI units

In 2018 a fundamental change was made to the SI at the 26th General Conference, this change being adopted on 20 May 2019: Seven "defining constants" now form the foundation of the SI and therefore provide the basis for the definition of the base units and all further SI units (source: [3]).

Second (time)

The second, symbol s, is the SI unit of time. It is defined by taking the fixed numerical value of the cesium frequency $\Delta\nu_{Cs}$, the unperturbed ground-state hyperfine-transition frequency of the cesium-133 atom to be 9 192 631 770 when expressed in the unit Hz, which is equal to s^{-1}.

In other words, one second is equal to the duration of 9 192 631 770 periods of the radiation corresponding to the energy of the transition between the two hyperfine levels of the unperturbed ground state of the ^{133}Cs atom.

Meter (length)

The meter, symbol m, is the SI unit of length. It is defined by taking the fixed numerical value of the speed of light in a vacuum c to be 299 792 458 when expressed in the unit m/s, where the second is defined in terms of $\Delta\nu_{Cs}$.

In other words, one meter is the length of the path traveled by light in a vacuum during a time interval with duration of 1/299 792 458 of a second.

Kilogram (mass)
The kilogram, symbol kg, is the SI unit of mass. It is defined by taking the fixed numerical value of the Planck constant h to be $6.62607015 \times 10^{-34}$ when expressed in the unit J s, which is equal to kg m^2 s^{-1}, where the meter and the second are defined in terms of c and $\Delta\nu_{Cs}$.

In other words, the unit kg is linked with the action, a quantity in theoretical physics with the unit kg m^2 s^{-1}. Together with the definitions of the second and the meter this leads to a definition of the kilogram as a function of the Planck constant h.

Ampere (electric current)
The ampere, symbol A, is the SI unit of electric current. It is defined by taking the fixed numerical value of the elementary charge e to be $1.602176634 \times 10^{-19}$ when expressed in the unit C, which is equal to A s, where the second is defined in terms of $\Delta\nu_{Cs}$.

In other words, an ampere is the current corresponding to the flow of $1 / (1.602176634 \times 10^{-19})$ elementary charges (electrons) per second.

Kelvin (thermodynamic temperature)
The kelvin, symbol K, is the SI unit of thermodynamic temperature. It is defined by taking the fixed numerical value of the Boltzmann constant k to be 1.380649×10^{-23} when expressed in the unit J K^{-1}, which is equal to kg m^2 s^{-2} K^{-1}, where the kilogram, the meter and the second are defined in terms of h, c and $\Delta\nu_{Cs}$.

In other words, one kelvin is equal to the change of thermodynamic temperature that results in a change of thermal energy (kT) by $1.380649 \cdot 10^{-23}$ J.

Mole (amount of substance)
The mole, symbol mol, is the SI unit of amount of substance. One mole contains exactly $6.02214076 \times 10^{23}$ elementary entities. This number is the fixed numerical value of the Avogadro constant N_A when expressed in the unit mol^{-1} and is called the Avogadro number. The amount of substance, symbol n, of a system is a measure of the number of specified elementary entities. An elementary entity may be an atom, a molecule, an ion, an electron, any other particle or a group of such particles of precisely specified composition.

In other words, a mole is the amount of substance of a system that contains $6.02214076 \cdot 10^{23}$ specified elementary entities.

Candela (luminous intensity)
The candela, symbol cd, is the SI unit of luminous intensity in a given direction. It is defined by taking the fixed numerical value of the luminous efficacy K_{cd} of monochromatic radiation of frequency 540×10^{12} Hz to be 683 when expressed in the unit lm W^{-1}, which is equal to cd sr W^{-1} or cd sr kg^{-1} m^{-2} s^3, where the kilogram, the meter and the second are defined in terms of h, c and $\Delta\nu_{Cs}$.

In other words, one candela is the luminous intensity (in a given direction) of a source that emits monochromatic radiation of frequency 540×10^{12} Hz and has a radiant intensity in that direction of 1/683 W sr^{-1}.

Decimal fractions and multiples of SI units
Decimal fractions and multiples of SI units (base SI units and derived SI units) are denoted by prefixes before the name of the unit (e.g., milligram) or by prefix symbols before the unit symbol (e.g., mg) (Table 2). The prefix symbol is placed without a gap in front of the unit symbol to form a coherent unit.

Prefixes are not used before further units of angle (degree, min, second), of time (minute, hour), and of temperature (degree Celsius).

Table 2: Prefixes for unit of measurements in acc. with DIN 1301 [2]

Prefix	Prefix symbol	Factor	Name of factor
yokto	y	10^{-24}	septillionth
zepto	z	10^{-21}	sextillionth
atto	A	10^{-18}	trillionth
femto	F	10^{-15}	thousand billionth
pico	p	10^{-12}	billionth
nano	N	10^{-9}	thousand millionth
micro	µ	10^{-6}	millionth
milli	m	10^{-3}	thousandth
centi	C	10^{-2}	hundredth
deci	d	10^{-1}	tenth
deca	da	10^{1}	ten
hecto	H	10^{2}	hundred
kilo	K	10^{3}	thousand
mega	m	10^{6}	million
giga	G	10^{9}	thousand million[1]
tera	T	10^{12}	billion[1]
peta	p	10^{15}	thousand billion
exa	E	10^{18}	trillion
zetta	z	10^{21}	sextillion
yotta	y	10^{24}	septillion

[1] In the USA: 10^9 = 1 billion, 10^{12} = 1 trillion.

Derived SI units

SI units are the seven base SI units and all the units derived from them, i.e., which can be represented as a product of powers of the base units. Thus, for example, the unit of force is obtained from Newton's Law $F = m\,a$ as

$$1\ \text{kg}\ \frac{\text{m}}{\text{s}^2} = 1\ \text{N (Newton)}.$$

If the power product contains only the factor 1, the units are referred to as coherently derived units. There are a total of 22 coherently derived units, which like the newton have been assigned their own names (Table 3).

Table 3: Derived SI units with special names

Quantity	Unit	Unit symbol	Expressed in other SI units
Plane angle	radian	rad	1
Solid angle	steradian	sr	1
Frequency	hertz	Hz	1 Hz = 1/s
Force	newton	N	1 N = 1 kg m/s^2 = 1 J/m
Pressure	pascal	Pa	1 Pa = 1 N/ m^2
Energy, work, quantity of heat	joule	J	1 J = 1 N m = 1 W s
Power	watt	W	1 W = 1 J/s = 1 V A
Celsius temperature	degree Celsius	°C	
Voltage	volt	V	1 V = 1 W/A
Electrical conductance	siemens	s	1 S = 1 A/V = 1/Ω
Electrical resistance	ohm	Ω	1 Ω = 1 V/A
Electric charge	coulomb	C	1 C = 1 A s
Electrical capacitance	farad	F	1 F = 1 C/V
Inductance	henry	H	1 H = 1 Wb/A
Magnetic flux	weber	Wb	1 Wb = 1 V s
Magnetic flux density, induction	tesla	T	1 T = 1 Wb/m^2
Luminous flux	lumen	lm	1 lm = 1 cd sr
Luminous intensity	lux	lx	1 lx = 1 lm/m^2
Radioactivity	becquerel	Bq	1 Bq = 1/s
Absorbed dose	gray	Gy	1 Gy = 1 J/kg
Dose equivalent	sievert	Sv	1 Sv = 1 J/kg
Catalytic activity	katal	kat	1 kat = 1 mol/s

Legal units

The German Units in Metrology and Determination of Time Act (EinhZeitG), rewritten on 22 February 1985, last amended on 18 July 2016 (as at 2021), stipulates the use of legal units in business and official transactions in Germany. Legal units are:

- the SI units,
- decimal fractions and multiples of SI units,
- other permitted units; see the overview on the following pages and [3].

The following tables provide an overview in accordance with DIN 1301 [2].

Table 4: Legal units

Quantity and symbol	Legal units SI	Others	Name	Relationship	Remarks and units not to be used, incl. their conversion
1. Length, area, volume.					
Length l	m		meter		
Area A	m^2		square meter		
		A	are	1 a = 100 m^2	
		ha	hectare	1 ha = 100 a = 10^4 m^2	
Volume V	m^3		cubic meter		
		l, L	liter	1 l = 1 L = 1 dm^3 = 10^3 cm^3 = 10^{-3} m^3	
2. Angle.					
(Plane) angle[1] α, β etc.	rad		radian	1 rad = $\frac{1\text{ m ar}}{1\text{ m radius}}$	1^g (centesimal degree) = 1 gon; 1^c (centesimal minute) = 10^{-2} gon; 1^{cc} (centesimal second) = 10^{-4} gon
		°	degree	1° = $\frac{\pi}{180}$ rad	
		′	minute	1° = 60′ = 3,600″	
		″	second		
	gon	gon	gon	1 gon = $\frac{\pi}{200}$ rad	
Solid angle[1] Ω	sr		steradian	1 sr = $\frac{1\,m^2\text{ spherical surface}}{1\,m^2\text{ sphere radius}^2}$	

[1] The units rad and sr can be replaced by the numeral 1 in calculations.

Quantity and symbol	Legal units SI	Others	Name	Relationship	Remarks and units not to be used, incl. their conversion
3. Mass.					
Mass m (weight)[1]	kg		kilogram		
		G	gram	1 g = 10^{-3} kg	
		T	metric ton	1 t = 10^3 kg	
Density ρ	kg/m^3			1 kg/m^3 = 1 g/dm^3	Weight of unit volume γ (kp/dm^3 or p/cm^3) $\gamma = \rho\, g$
Moment of inertia J (mass moment, 2nd order)	kg m^2			$J = m\, r^2$ r = radius of gyration	Flywheel effect $G\, D^2$ in kp m^2 $D = 2r$, $G = m\, g$ $G\, D^2 = 4 J\, g$
4. Time quantities.					
Time t Duration Interval[2]	s		second		In energy management one year is calculated as 8,760 hours
		min	minute	1 min = 60 s	
		h	hour	1 h = 60 min	
		d	day	1 d = 24 h	
		a	year		
Frequency f	Hz		hertz	1 Hz = 1/s	
Rotational speed (rotational frequency) n	s^{-1}			1 s^{-1} = 1/s	rpm and r/min (revolutions per minute) are still permissible for expressing rotational speed, but is better replaced by min^{-1} (1 rpm = 1 r/min = 1 min^{-1})
		min^{-1} 1/min		1 min^{-1} = 1/min = (1/60) s^{-1}	
Angular frequency ω	s^{-1}			$\omega = 2\pi f$	
Velocity v	m/s	km/h		1 km/h = (1/3.6) m/s	
Acceleration[4] a	m/s^2			Normal acceleration of free fall $g \approx 9.80665$ m/s^2	
Angular velocity[3] ω	rad/s				
Angular acceleration[3] α	rad/s^2				

[1] The term "weight" is ambiguous in everyday usage; it is used to denote mass as well as weight (DIN 1305 [4]).

[2] Clock times: h, min, s written as superscripts. Example: $3^h\ 25^m\ 6^s$.

[3] The unit rad can be replaced by the numeral 1 in calculations.

[4] Acceleration is sometimes expressed in m/s^2 as a multiple of gravitational acceleration g.

Quantity and symbol	Legal units SI	Others	Name	Relationship	Remarks and units not to be used, incl. their conversion
5. Force, energy, power.					
Force F Force due to weight G	N N		newton	1 N = 1 kg m/s^2	1 kp (kilopond) ≈ 9.80665 N
Impulse p	N s			1 N s = 1 kg m/s	
Pressure, gen. p	Pa		pascal	1 Pa = 1 N/m^2	1 at (techn. atmosphere) = 1 kp/cm^2 ≈ 0.980665 bar p ≈ 1.01325 bar ≈ 1,013.25 hPa (standard value of air pressure)
		bar	bar	1 bar = 10^5 Pa	
Mechanical stress σ, τ	N/m^2			1 N/m^2 = 1 Pa	1 kp/m^2 ≈ 9.80665 N/m^2
		N/mm^2		1 N/mm^2 = 1 MPa	
Hardness	Brinell and Vickers hardness are no longer given in kp/mm^2. Instead, an abbreviation of the relevant hardness is written as the unit after the numerical value used previously (including an indication of the test force, etc., where applicable).				Examples: Previously Now HB = 350 kp/mm^2 350 HB HV30 = 720 kp/mm^2 720 HV30 HRC = 60 60 HRC
Energy E Work W	J		joule	1 J = 1 N m = 1 W s = 1 kg m^2/s^2	1 kcal (kilocalorie) = 4.1868 kJ
Heat, quantity of heat[1] Q		W s	watt-second	1 W s = 1 J	
		kW h	kilowatt-hour	1 kW h = 3.6 MJ	
		eV	electron-volt	1 eV = 1.602176634 · 10^{-19} J	
Torque M	N m		newton meter		1 kp m (kilopondmeter) ≈ 9.80665 N m
Power P Heat flow $\dot{Q}$ Radiated power Φ	W		watt	1 W = 1 J/s = 1 N m/s	1 kp m/s ≈ 9.80665 W 1 hp (horsepower) ≈ 0.7355 kW
Apparent power P_s	VA		volt-ampere	1 VA = 1 W	
Reactive power P_q	var		var	1 var = 1 W	

[1] The quantity of heat is expressed in joules.

Quantity and symbol	Legal units SI	Others	Name	Relationship	Remarks and units not to be used, incl. their conversion
6. Viscosimetric quantities.					
Dynamic viscosity η	Pa s		pascal-second	1 Pa s = 1 N s/m^2 = 1 kg/(s m)	1 P (poise) = 0.1 Pa s
Kinematic viscosity ν	m^2/s			1 m^2/s = 1 Pa s/(kg/m^3)	1 St (stokes) = 10^{-4} m^2/s
7. Temperature and heat.					
Temperature T	K	kelvin		$\vartheta = (T - 273.15\ \mathrm{K})\,\frac{°C}{K}$	
Temperature ϑ		°C	degree Celsius		
Temperature difference ΔT	K		kelvin	1 K = 1 °C	In the case of composite units, express temperature differences in K
Temperature difference $\Delta\vartheta$		°C	degree Celsius		
For quantity of heat and heat flow, refer to 5					
Specific heat capacity (spec. heat) c	J/(kg K)				1 kcal/(kg K) = 4.1868 kJ/(kg K)
Molar heat capacity C	J/(mol K)				
Thermal conductivity λ	W/(m K)				1 kcal/(m h K) = 1.163 W/(m K)
8. Electrical quantities.					
Electric current I	A		ampere		
Electric potential (voltage) U	V		volt	1 V = 1 W/A	
Electrical conductance: Conductance G Susceptance B Admittance Y	s		siemens	1 S = 1 A/V = 1/Ω	
Electrical resistance: Resistance R Reactance X Impedance Z	Ω		ohm	1 Ω = 1/S = 1 V/A	
Quantity of electricity, electric charge Q	C		coulomb	1 C = 1 A s	
		A h	ampere-hour	1 A h = 3,600 C	
Electrical capacitance C	F		farad	1 F = 1 C/V	

Quantity and symbol	Legal units SI	Others	Name	Relationship	Remarks and units not to be used, incl. their conversion
Electrical flux density, displacement D	C/m^2			$1\ C/m^2 = 1\ A\,s/m^2$	
Electric field strength E	V/m			$1\ V/m = 1\ W/(A\,m)$	

9. Magnetic quantities.

Inductance L	H		henry	$1\ H = 1\ Wb/A$	
Magnetic flux Φ	Wb		weber	$1\ Wb = 1\ V\,s$	$1\ M\ (maxwell) = 10^{-8}\ Wb$
Magnetic flux density, induction B	T		tesla	$1\ T = 1\ Wb/m^2$	$1\ G\ (gauss) = 10^{-4}\ T$
Magnetic field strength H	A/m			$1\ A/m = 1\ N/Wb$	$1\ Oe\ (oersted) = \frac{10^3}{(4\pi)}\ \frac{A}{m}$

10. Photometric quantities.

Luminous intensity I	cd		candela		
Luminance L	cd/m^2				$1\ sb\ (stilb) = 10^4\ cd/m^2$ $1\ asb\ (apostilb) = 1/\pi\ cd/m^2$
Luminous flux Φ	lm		lumen	$1\ lm = 1\ cd\ sr$ (sr = steradian)	
Illuminance E	lx		lux	$1\ lx = 1\ lm/m^2$	

11. Acoustic quantities.

Sound pressure p	Pa		pascal		
Sound intensity I	W/m^2				
Sound pressure level L_p	Np		neper	$1\ Np = 1$ (nondimensional)	$L_p = \ln(p_1/p_2)\ Np$ $= L_I = 0.5 \ln(I_1/I_2)\ Np$
Sound intensity level L_I		dB	decibel	$1\ dB = \frac{1}{20} \ln 10\ Np \approx 0.1151\ Np$	$L_p = 20 \lg(p_1/p_2)\ dB$ $= L_I = 10 \lg(I_1/I_2)\ dB$ At $f = 1{,}000$ Hz the (physiological) loudness level is measured in "phon": 1 phon = 1 dB and the loudness in "sone": 1 sone = 40 phon
		B	bel	$1\ B = 10\ dB \approx 1.151\ Np$	

Quantity and symbol	Legal units SI	Others	Name	Relationship	Remarks and units not to be used, incl. their conversion
Sound pressure level L_{pA}		dB (A)			Frequency-dependent weighting tuned to the human ear at 20 to 40 phon
Sound power level A-weighted L_{WA}					

12. Quantities used in atom physics and other quantities.

Energy W		eV	electron-volt	1 eV = $1.602176634 \cdot 10^{-19}$ J	
Activity of a radioactive substance A	Bq		becquerel	1 Bq = 1 s^{-1}	1 Ci (curie) = $3.7 \cdot 10^{10}$ Bq
Absorbed dose D	Gy		gray	1 Gy = 1 J/kg	1 rd (rad) = 10^{-2} Gy
Dose equivalent D_q	Sv		sievert	1 Sv = 1 J/kg	1 rem = 10^{-2} Sv
Absorbed dose rate $\dot{D}$	Gy/s			1 Gy/s = 1 W/kg	1 rd/s = 10^{-2} Gy/s
Ion dose J	C/kg				1 R (roentgen) = $258 \cdot 10^{-6}$ C/kg
Ion dose rate $\dot{J}$	A/kg				
Amount of substance n	mole		mole		
Catalytic activity	kat		katal	1 kat = 1 mol/s	

Further units

Table 5: Conversion of units

Units of length

Name	*Conversion*
micron	1 µ = 1 µm
typographical point	1 p = 0.376065 mm
inch	1 in = 25.4 mm
foot	1 ft = 12 in = 0.3048 m
yard	1 yd = 3 ft = 0.9144 m
mile	1 mile = 1,760 yd = 1.609344 km
nautical mile (international mile)	1 NM = 1 sm = 1.852 km (≈ 1′ of degree of longitude)

Anglo-American units of length:

microinch	1 µin = 0.0254 µm
milliinch	1 mil = 0.0254 mm
link	1 link = 201.17 mm
rod	1 rod = 1 pole = 1 perch = 5.5 yd = 5.0292 m
fathom	1 fathom = 2 yd = 1.8288 m
chain	1 chain = 22 yd = 20.1168 m
furlong	1 furlong = 220 yd = 201.168 m

Units of area

Name	*Conversion*
square inch (sq in)	1 in^2 = 6.4516 cm^2
square foot (sq ft)	1 ft^2 = 144 in^2 ≈ 0.0929 m^2
square yard (sq yd)	1 yd^2 = 9 ft^2 ≈ 0.8361 m^2
acre	1 ac = 4840 yd^2 ≈ 4,046.856 m^2
square mile (sq mile)	1 $mile^2$ = 640 acre ≈ 2.59 km^2
barn	1 b = 10^{-28} m^2

Units of volume

Name	*Conversion*
cubic inch (cu in)	1 in^3 = 16.3871 cm^3
cubic foot (cu ft)	1 ft^3 = 1,728 in^3 = 0.02832 m^3
cubic yard (cu yd)	1 yd^3 = 27 ft^3 = 0.76456 m^3

Further units in the United Kingdom (UK):

fluid ounce	1 fl oz ≈ 0.028413 *l*
pint	1 pt ≈ 0.56826 *l*
quart	1 qt = 2 pt ≈ 1.13652 *l*
gallon	1 gal = 4 qt ≈ 4.5461 *l*
barrel (crude oil)	1 bbl = 35 gal ≈ 159.1 *l*
barrel (other liquids)	1 bbl = 36 gal ≈ 163.6 *l*

Further units in the United States (US):

fluid ounce	1 fl oz ≈ 0.029574 *l*
liquid pint	1 liq pt ≈ 0.47318 *l*
liquid quart	1 liq qt = 2 liq pt ≈ 0.94635 *l*
gallon	1 gal = 4 liq qt ≈ 3.7854 *l*
liquid barrel	1 liq bbl = 31,5 gal ≈119.24 *l*
barrel petroleum	1 barrel petroleum = 42 gal ≈ 158.99 *l* (for crude oil)

Velocities

Name	*Conversion*
kilometer per hour	1 km/h = (1/3.6) m/s ≈ 0.2778 m/s
miles per hour	1 mile/h ≈ 1.6093 km/h
knot	1 kn = 1 NM/h = 1 sm/h = 1.852 km/h ≈ 0.5144 m/s
Mach number Ma	The Mach number Ma is the quotient of velocity and velocity of sound.

Units of mass

Name	*Conversion*
Pfund (pound)	1 Pfund = 0.5 kg
Zentner (metric hundredweight)	1 Ztr = 50 kg
Doppelzentner (metric quintal)	1 dz = 100 kg
Gamma	1 γ = 1 µg
metric carat (for precious stones only)	1 Kt = 0.2 g
atomic unit of mass	1 u ≈ 1.660539·10^{-27} kg
grain	1 gr = 64.79891 mg
pennyweight	1 dwt = 24 gr ≈ 1.5552 g
dram	1 dr ≈ 1.77184 g
ounce	1 oz = 16 dram ≈ 28.3495 g
troy ounce	1 oz tr (US) = 1 oz tr (UK) ≈ 31.1035 g
pound	1 lb = 16 oz ≈ 453.592 g
stone (UK)	1 st = 14 lb ≈ 6.35 kg
quarter (UK)	1 qr = 28 lb ≈ 12.7 kg
slug (mass accelerated by 1 ft/s^2 when 1 lbf is exerted on it)	1 slug ≈ 14.4939 kg
hundredweight (US)	1 cwt = 1 cwt sh (short cwt) = 1 quintal = 100 lb ≈ 45.3592 kg
hundredweight (UK)	1 cwt = 1 cwt l (long cwt) = 112 lb ≈ 50.8023 kg
ton (US)	1 ton (US) = 1 tn sh (short ton) ≈ 0.90718 t
ton (UK)	1 ton (UK) = 1 tn l (long ton) = 1 ton dw (ton deadweight) ≈ 1.01605 t
–	1 t dw = 1 t

Density

Name	*Conversion*
	1 lb/ft^3 = 16.018 kg/m^3
	1 lb/gal (UK) = 99.776 kg/m^3
	1 lb/gal (US) = 119.83 kg/m^3

Areometer degrees n are a measure of the density ρ of a liquid relative to the density of water at 15 °C.
$n = 144.3(\rho - 1\ kg/l)/\rho$ °Bé (degree Baumé)
$n = (141.5\ kg/l - 131.5 \cdot \rho)/\rho$ °API
(API: American Petroleum Institute)

Units of force

Name	*Conversion*
pond	1 p ≈ 9.80665 mN
kilopond	1 kp ≈ 9.80665 N
dyne	1 dyn = 10^{-5} N
sthène (French)	1 sn = 1 kN
poundforce	1 lbf = 4.44822 N
poundal (force which accelerates a mass of 1 lb by 1 ft/s^2)	1 pdl = 0.138255 N

Units of pressure and stress

Name	*Conversion*
microbar	1 µbar = 0.1 Pa
millibar	1 mbar = 1 hPa = 100 Pa
bar	1 bar = 10^5 Pa
–	1 kp/mm^2 ≈ 9.80665 x 10^6 Pa
technical atmosphere	1 at = 1 kp/cm^2 ≈ 9.80665 · 10^4 Pa
physical atmosphere	1 atm = 1.01325 · 10^5 Pa
torr	1 torr = 1 mm Hg (mercury column) = 133.322 Pa
–	1 mm water column = 1 kp/m^2 ≈ 9.80665 Pa

The following units are derived from the technical atmosphere:
p_{abs} (absolute pressure): 1 ata
p_{amb} (ambient pressure)
$p_e = p_{abs} - p_{amb}$ (gage pressure): 1 atpaa for $p_{abs} > p_{amb}$
$p_e = p_{amb} - p_{abs}$ (vacuum): 1 atpba for $p_{abs} < p_{amb}$

Anglo-American and other units:	
poundforce per square inch	1 lbf/in^2 = 1 psi = 6894.76 Pa
poundforce per square foot	1 lbf/ft^2 = 1 psf = 47.8803 Pa
tonforce per square inch (UK)	1 tonf/in^2 = $1.54443 \cdot 10^7$ Pa
poundal per square foot	1 pdl/ft^2 = 1.48816 Pa
barye (French)	1 barye = 0.1 Pa
pièce (French)	1 pz = 1 sn/m^2 (sthène/m^2) = 1,000 Pa

Units of energy

Name	*Conversion*
erg	1 erg = 10^{-7} J
calorie[1]	1 cal = 4.1868 J
	1 kp · m = 9.80665 J
	1 HP · h = $2.6478 \cdot 10^6$ J
therm	1 therm = $105.50 \cdot 10^6$ J

Anglo-American and other units:	
inch ounceforce	1 in ozf = 7.062 mJ
foot poundal	1 ft pdl = 0.04214 J
inch poundforce	1 in lbf = 0.11299 J
foot poundforce	1 ft lbf = 1.35582 J
British thermal unit[2]	1 Btu = 1,055.06 J
therm	1 therm = 10^5 Btu
horsepower hour	1 hp · h = $2.685 \cdot 10^6$ J
thermie (French)	1 thermie = 1,000 frigories = 1,000 kcal = 41.868 MJ
kg coal equivalents	1 kg CE = 29.3076 MJ (for the calorific value H_i = 7,000 kcal/kg coal)
tonne of coal equivalents	1 t CE = 1,000 kg CE

Units of power

Name	*Conversion*
kilocalorie per hour	1 kcal/h = 1.163 W
calorie per second	1 cal/s = 4.1868 W
–	1 kp · m/s = 9.80665 W
horsepower, cheval vapeur (French)	1 HP = 1 ch = 735.499 W

Anglo-American units:

–	1 ft · lbf/s = 1.35582 W
horsepower	1 hp = 745.70 W
–	1 Btu/s = 1,055.06 W

Fuel consumption

Name	*Conversion*
–	1 g/(HP · h) = 1.3596 g/(kW h)
–	1 lb/(hp · h) = 608.277 g/(kW h)
–	1 liq pt/(hp · h) = 634.545 cm^3/(kW h)

x mile/gal (US) $\triangleq \frac{235.21}{x}$ *l*/100 km

x mile/gal (UK) $\triangleq \frac{282.48}{x}$ *l*/100 km

y *l*/100 km $\triangleq \frac{235.21}{y}$ mile/gal (US)

y *l*/100 km $\triangleq \frac{282.48}{y}$ mile/gal (UK)

[1] The quantity of heat which is required to heat 1 g water from 15 °C to 16 °C.
[2] The quantity of heat which is required to heat 1 lb water from 63 °F to 64 °F.

Units of temperature

Name	*Conversion*
K (kelvin)	
°C (degree Celsius)	$\vartheta/°C = T/K - 273.15$ This scale is based on the freezing point of water at 0 °C and the boiling point at 100 °C, at normal pressure in each case
°F (degree Fahrenheit)	$T_F/°F = 1.8 \cdot T/K - 459.67$ The freezing point of water is 32 °F, human body temperature is 96 °F
°Ra (degree Rankine)	$T_{Ra}/°Ra = 1.8 \cdot T/K$ This scale begins with the absolute temperature zero point at 0 K and uses a graduation like the Fahrenheit
°Re (degree Réaumur)	$T_{Re}/°Re = 0.8 \cdot (T/K - 273.15)$ The freezing point of water is 0 °Re, the boiling point is 80 °Re
Temperature differences	1 K ≙ 1 °C ≙ 1.8 °F ≙ 1.8 °Ra ≙ 0.8 °Re

Units of viscosity (kinematic)

Name	*Conversion*
	1 ft^2/s = 0.092903 m^2/s Measurement is carried out by viscosimeter in accordance with the standard DIN EN ISO 2431 [5]
A-seconds	Runout time from flow cup
Engler number	Relative runout time from Engler device: 1 °E ≈ 7.6 mm^2/s
RI seconds	Runout time from Redwood-I viscosimeter (UK): 1 R″ ≈ 4.06 mm^2/s
SU seconds	Runout time from Saybolt Universal viscosimeter (US): 1 S″ ≈ 4.63 mm^2/s

Natural constants

It is currently assumed that the physical limits apply equally at any place in our world. These describe the natural constants which both contain physical quantities (e.g., speed of light in vacuum, masses of elementary particles) and represent the connections between them (e.g., field constants of gravitation, electricity and magnetism). Their values depend on the system of units and must be experimentally determined.

Since the 26th CGPM seven constants have formed the foundation of the SI international system of units. These constants are used to define the base units and for this reason are referred to as defining constants (Table 6). The numerical values of these constants can thus be specified precisely.

Table 6: Defining constants

Hyperfine frequency of the cesium atom	$\Delta\nu_{Cs}$	9 192 631 770 Hz
Speed of light (in vacuum)	c	299 792 458 $\frac{m}{s}$
Planck constant	h	$6.626\,070\,15 \cdot 10^{-34}$ J s
Elementary charge	e	$1.602\,176\,634 \cdot 10^{-19}$ C
Boltzmann constant	k	$1.380\,649 \cdot 10^{-23}$ $\frac{J}{K}$
Avogadro constant	N_A	$6.022\,14076 \cdot 10^{23}$ mol^{-1}
Luminous efficacy of radiation		683 lm W^{-1}

Table 7: Selection of derived, exact constants

Faraday constant	F	$F = e\,N_A$ = 96 485.333 12… $\frac{C}{mol}$
Universal gas constant	R	$R = k\,N_A$ = 8.314 462 6318… $\frac{J}{mol \cdot K}$
Stefan-Boltzmann constant	σ	$5.670\,374\,419\ldots \cdot 10^{-8}$ $\frac{W}{m^2 \cdot K^4}$

Table 8: Further natural constants

Loschmidt number	N_L	$N_L = \frac{N_A}{V_{m0}}$ $\approx 2.6867805(24) \cdot 10^{25}\ \mathrm{m^{-3}}$ V_{m0}: Molar volume of an ideal gas under normal conditions
electric field constant	ε_0	$\frac{1}{\mu_0\ c^2}$ $\approx 8.854\,187\,817\,62 \cdot 10^{-12}\ \frac{\mathrm{F}}{\mathrm{m}}$
magnetic field constant	μ_0	$4\pi \cdot 10^{-7}\ \frac{\mathrm{V\,s}}{\mathrm{A\,m}}$ $\approx 12.566\,370\,614 \cdot 10^{-7}\ \frac{\mathrm{H}}{\mathrm{m}}$
gravitation constant	G	$6.674\,2(10) \cdot 10^{-11}\ \frac{\mathrm{m^3}}{\mathrm{kg \cdot s^2}}$
atomic unit of mass	u	$1.660\,538\,86(28) \cdot 10^{-27}$ kg
rest mass of electron	m_e	$9.109\,382\,6(16) \cdot 10^{-31}$ kg
rest mass of proton	m_p	$1.672\,621\,71(29) \cdot 10^{-27}$ kg

Constants derived from these defining constants can also be specified precisely (Table 7).

Table 8 lists further important natural constants (source: [3]). The digits in parentheses after a numerical value denote the uncertainty in the last places of the value.

References

[1] ISO 80000: Quantities and units (2013).
[2] DIN 1301: Units – Part 1: Unit names, unit symbols (2010).
[3] Die gesetzlichen Einheiten in Deutschland. 2nd Edition, Physikalisch-Technische Bundesanstalt Nationales Metrologieinstitut, 2019, Braunschweig.
[4] DIN 1305: Mass, as weighed value, force, weight force, weight, load; concepts (1988).
[5] DIN EN ISO 2431: Paints and varnishes – Determination of flow time by use of flow cups (ISO 2431:2011); German version EN ISO 2431:2011.

Mathematical signs and symbols

+	plus
–	minus
· or ×	multiplied by
: or /	divided by
=	equal to
≈	approximately equal to
≠	not equal to
<	less than
>	greater than
≤	less than or equal to
≥	greater than or equal to
~ or ∝	proportional to
$\sum a_i$	sum over a_i
$\prod a_i$	product of a_i
$n!$	n factorial $(1 \cdot 2 \cdot 3 \ldots n)$
Δ	difference or Laplace operator
$\sqrt{\ }$	root
∥	parallel to
⊥	perpendicular to
→	approaches
∞	infinity
d/dx	differentiation to x
$\partial/\partial x$	partial differentiation to x
$\int f(x)dx$	integral of $f(x)$

Greek alphabet

A	α	Alpha
B	β	Beta
Γ	γ	Gamma
Δ	δ	Delta
E	ε, ϵ	Epsilon
Z	ζ	Zeta
H	η	Eta
Θ	θ, ϑ	Theta
I	ι	Iota
K	κ, $\varkappa$	Kappa
Λ	λ	Lambda
M	μ	Mu
N	ν	Nu
Ξ	ξ	Xi
O	o	Omicron
Π	π, ϖ	Pi
P	ρ, ϱ	Rho
Σ	σ, ς	Sigma
T	τ	Tau
Y, Υ	υ	Upsilon
Φ	ϕ, φ	Phi
X	χ	Chi
Ψ	ψ	Psi
Ω	ω	Omega

Basic principles of mechanics

Rectilinear and rotary motion

Forces, moments and energies are required to move bodies with mass. Any number of motions can be composed of rectilinear motions (translation) and rotary motions (rotation). The most important basic equations for these motions are set out in Table 2. The quantities used are listed in Table 1.

Mass and moment of inertia

Mass m is a property of matter and the cause of the inertia which sets a resistance against a change of velocity (acceleration) in the case of a rectilinear motion. The moment of inertia J (also called mass moment of inertia or rotating mass) causes a resistance in the case of a rotary motion (Table 3).

Path, velocity and acceleration

Path s is a limited distance, velocity v is the path covered during a specific time t. For a rotary motion angular velocity ω is obtained from the angle φ covered during a specific time. A uniform motion exists when velocity v or rotational speed n (or angular velocity ω) is constant. In this case, acceleration (a or α) equals zero.

A body is accelerated when there is a change of velocity. A motion is uniformly accelerated if acceleration is constant. In the case of negative acceleration, the motion is decelerated or braked.

Force and moment

A force accelerates or deforms a body. It is referred to in classic physics as the time rate of change of linear momentum p. If a mass m is rotated at centroidal dis-

Table 1: Symbols and units

Quantity		Unit
A	Area	m^2
E	Modulus of elasticity	N/mm^2
E_k	Kinetic energy	J = N m
E_p	Potential energy	J = N m
E_{rot}	Rotational energy	J = N m
F	Force	N
F_G	Weight	N
F_m	Mean force during impulse period	N
F_R	Frictional force	N
F_U	Peripheral force	N
F_Z	Centrifugal force	N
H	Rotational impulse	$N m \cdot s = kg \cdot m^2/s$
I	Force impulse	$N s = kg \cdot m/s$
J	Moment of inertia	$kg \cdot m^2$
L	Angular momentum	$N m \cdot s$
M_t	Torque	N m
$M_{t,m}$	Mean torque during impulse period	N m
P	Power	W = N m/s
R_e	Yield point	N/mm^2
R_m	Tensile strength	N/mm^2
V	Volume	m^3
W	Work, energy	J = N m
a	Acceleration	m/s^2

Quantity		Unit
a_Z	Centrifugal acceleration	m/s^2
d	Diameter	m
e	Base of natural logarithms ($e \approx 2.781$)	–
g	Acceleration of free fall ($g \approx 9.81$)	m/s^2
h	Height	m
i	Radius of gyration	m
l	Length	m
m	Mass	kg
n	Rotational speed	1/s
p	Linear momentum	N s
p	Surface pressure	N/mm^2
r	Radius	m
s	Length of path	m
t	Time	s
v	Velocity	m/s
α	Angular acceleration	rad/s^2
β	Wrap angle	°
γ	Wedge angle	°
ε	Elongation	%
μ	Coefficient of friction	–
ν	Transverse contraction	–
ρ	Density	kg/m^3
φ	Angle of rotation	rad
ω	Angular velocity	1/s

tance r about a rotational axis at angular velocity ω, this generates a centrifugal force F_Z which acts radially outwards from the center. A force which is applied at distance r from a center of rotation generates a moment. An accelerated rotary motion with rotating mass J causes a moment in the form of an acceleration or braking torque.

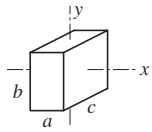
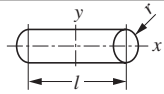
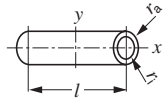
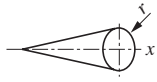
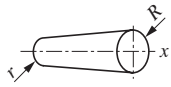
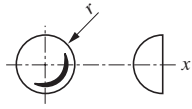
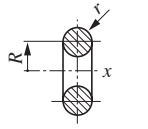
Work and energy

When a force F moves a body by path s, work W is performed which is stored as energy in this body. Energy is therefore defined in physics as stored work. Conversely, an energy can perform work. A distinction is made in mechanics between kinetic and potential energy. Kinetic energy is the work that must be expended to accelerate a body with mass m or rotating mass J to velocity v or angular velocity ω. Potential energy is the work that must be expended to raise a body to a height h. When a spring is tensioned, potential energy is stored which can perform work again when the spring is released.

Table 2: Rectilinear and rotary motion

Rectilinear motion (translation)		Rotary motion (rotation)	
Mass		**Moment of inertia** (Table 3)	
$m = V\rho$		$J = m\,i^2$	
Path		**Angle**	
$s = \int v(t)\,dt$		$\varphi = \int \omega(t)\,dt$	
$s = v\,t$	[v = const.]	$\varphi = \omega\,t = 2\pi n$	[n = const.]
$s = {}^1\!/_2\,a\,t^2$	[a = const.]	$\varphi = {}^1\!/_2\,\alpha\,t^2$	[α = const.]
Velocity		**Angular velocity**	
$v = ds(t)/dt$		$\omega = d\varphi(t)/dt$	
$v = s/t$	[v = const.]	$\omega = \varphi/t = 2\pi n$	[n = const.]
$v = a\,t = \sqrt{2\,a\,s}$	[a = const.]	$\omega = \alpha\,t = \sqrt{2\,\alpha\,\varphi}$	[α = const.]
		Peripheral velocity	
		$v = r\,\omega$	
Acceleration		**Angular acceleration**	
$a = dv(t)/dt$		$\alpha = d\omega(t)/dt$	
$a = (v_2 - v_1)/t$	[α = const.]	$\alpha = (\omega_2 - \omega_1)/t$	[α = const.]
		Centrifugal acceleration	
		$a_Z = r\,\omega^2$	
Force		**Torque**	
$F = m\,a$		$M_t = F\,r = J\,\alpha$	
		Centrifugal force	
		$F_Z = m\,r\,\omega^2$	
Work		**Rotational work**	
$W = F\,s$		$W = M_t\,\varphi$	
Translation energy		**Rotational energy**	
$E_k = {}^1\!/_2\,m\,v^2$		$E_{rot} = {}^1\!/_2\,J\,\omega^2$	
Potential energy			
$E_p = F_G\,h$			
Power		**Power**	
$P = dW/dt = F\,v$		$P = dW/dt = M_t\,\omega = M_t \cdot 2\,\pi\,n$	
Linear momentum		**Angular momentum**	
$p = m\,v$		$L = J\,\omega = J \cdot 2\,\pi\,n$	
Force impulse		**Rotational impulse**	
$I = \Delta p = F_m\,(t_2 - t_1)$		$H = M_{t,\,m}\,(t_2 - t_1)$	

Table 3: Moments of inertia

Type of solid	Moments of inertia (J_x about the x-axis[1], J_y about the y-axis[1])
Right parallelepiped, cuboid 	$J_x = m\frac{b^2+c^2}{12}$, $J_y = m\frac{a^2+c^2}{12}$ $J_x = J_y = m\frac{a^2}{6}$ (cube with side length a)
Regular cylinder 	$J_x = m\frac{r^2}{2}$, $J_y = m\frac{r^2+l^2}{12}$
Hollow regular cylinder 	$J_x = m\frac{r_a^2+r_i^2}{2}$ $J_y = m\frac{r_a^2+r_i^2+\frac{l^2}{3}}{4}$
Circular cone 	$J_x = m\frac{3r^2}{10}$ $J_x = m\frac{r^2}{2}$ Envelope of cone (excluding end base)
Truncated circular cone 	$J_x = m\frac{3(R^5-r^5)}{10(R^3-r^3)}$ $J_x = m\frac{(R^2+r^2)}{2}$ Envelope of cone (excluding end faces)
Sphere and hemisphere 	$J_x = m\frac{2r^2}{5}$ $J_x = m\frac{2r^2}{3}$ Surface area of sphere
Hollow sphere r_a Outer sphere radius r_i Inner sphere radius	$J_x = m\frac{2(r_a^5-r_i^5)}{5(r_a^3-r_i^3)}$
Torus  UAN0007-2Y	$J_x = m(R^2+\frac{3}{4}r^2)$

[1] The moment of inertia for an axis parallel to the x-axis or y-axis at a distance a is: $J_A = J_x + m\,a^2$ or $J_A = J_y + m\,a^2$.

Conservation of energy
The law of conservation of energy states that the total energy in a closed system is constant. Energy can be neither created nor destroyed, but instead converted between different forms of energy (e.g. kinetic energy to thermal energy) or transmitted from one body to another.

Power
Power is the work performed during a specific period of time. Because every power transfer is associated with loss, the output power is always less than the input power. The ratio of output to input power is called efficiency η and is therefore always less than 1.

$$\eta = \frac{P_{out}}{P_{in}}.$$

Linear momentum and impulse
Linear momentum p describes the motion of a body with mass, and is calculated as the product of moving mass m and velocity v. Every moving body can transmit its linear momentum during an impulse process to another body, as happens for example in a collision between two vehicles. The force acting on a body gives rise to a change of linear momentum, which is called force impulse I.

During a rotary motion angular momentum L is obtained from the product of rotating mass J and angular velocity ω. A rotational impulse H is generated for example when two disks are jerkily coupled.

Conservation of momentum
The law of conservation of momentum states that the total momentum in a closed system is constant. It follows from this that the total momentum before and after an impulse must be equal.

Statics

Statics is the science of equilibrium on a rigid body. A body is in equilibrium when it is at rest or is moving uniformly or in a straight line. Equilibrium prevails when the sum of the applied forces and moments in all directions is equal to zero.

Plane system of forces
Forces are vectors which are determined by size and direction. They are geometrically (vectorially) added:

$$\vec{F}_{res} = \vec{F}_1 + \vec{F}_2 .$$

Two forces are composed with a parallelogram of forces or a triangle of forces (Figure 1a). If there are several forces, the resultant force F_{res} is determined by means of a polygon of forces (Figure 1b). When the polygon of forces is closed, the system of forces is in equilibrium.

Figure 1: Composition of forces
a) Parallelogram of forces and triangle of forces,
b) Closed polygon of forces (system in equilibrium).
F Force,
F_{res} Resultant force.

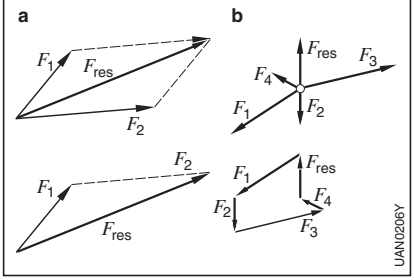
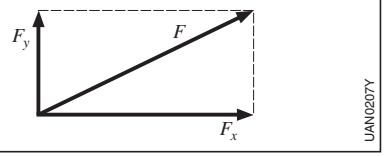


Figure 2. Force resolution
F Force,
F_x Force in x-direction,
F_y Force in y-direction.



A force can also be resolved into components. A resolution into orthogonal components is sensible (Figure 2).

Transmission of force
Mechanical machines for the transmission of force can be reduced to the "lever" and "wedge" principles.

Lever
The lever principle can be derived from the equilibrium condition "sum of the moments is equal to zero". Disregarding friction, the system in Figure 3 is in equilibrium for the following condition:

$M_{t1} = M_{t2}$ or $F_1 r_1 = F_2 r_2$.

Figure 3: Lever principle
F Force, r Radius.

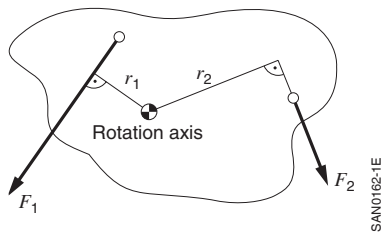
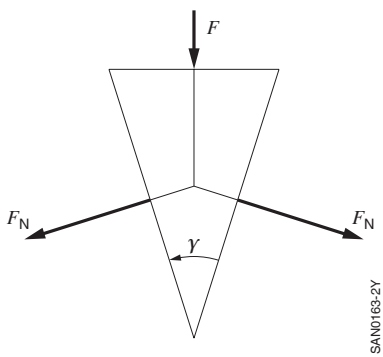


Figure 4: Forces on the wedge
F Insertion force, F_N Normal force,
γ Wedge angle.



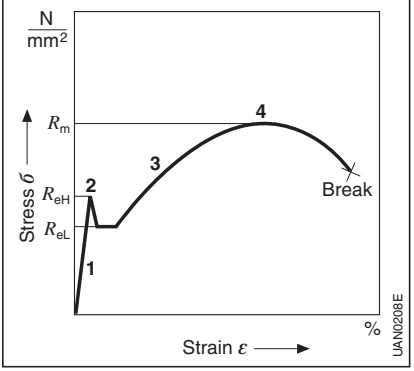
The lever principle is encountered in many applications, ranging from simple pliers, scales and wrenches through gear wheels and belt drives, right down to the connecting rods in piston engines.

Wedge
With the wedge principle, depending on the wedge angle γ, small forces (insertion force F) can be translated into large normal forces F_N (Figure 4). Without allowing for friction, the following applies:

$$F_N = \frac{F}{2 \sin \frac{\gamma}{2}} .$$

Very large application forces can be introduced in the smallest of spaces with the aid of wedges. Examples include wedges in shaft/hub connections and tapered connections for transmitting torques. However, both the screw and the clamping eccentric work according to the classic wedge principle.

Figure 5: Stress-strain diagram
1 Hooke's range,
2 Yield point,
3 Plastic deformation,
4 Tensile strength.
R_{eL} Lower yield point,
R_{eH} Upper yield point,
R_m Tensile strength
(maximum load before break).



Calculation of strength

Hooke's law

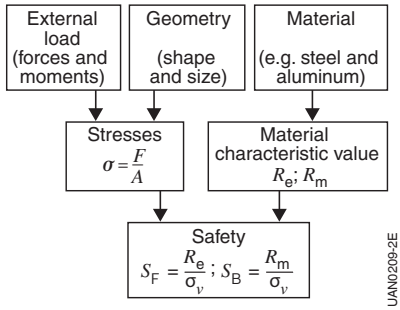
In response to an external load a body is deformed and stresses are generated inside the material (Figure 5). Metals have linear-elastic behavior until the yield point is reached. In other words, within this Hooke's range the component assumes its original length again when the load is reduced. The yield point R_e is the limit between elastic and plastic deformation. Above the yield point the material begins to "creep" and then remains permanently deformed. The tensile strength R_m is the maximum load before the component cracks or breaks.

In the case of ductile materials with a pronounced yield point R_e is always specified as the limit for dimensioning. The tensile strength is used for dimensioning only in the case of brittle materials which do not have a pronounced yield point (e.g. gray cast iron).

The linear area is known as Hooke's straight line, in which the ratio of stress σ and strain (elongation) ε is constant. With the proportionality constant E, which is termed the modulus of elasticity, Hooke's law for the mono-axial stress state is:

$$\sigma = E\varepsilon \ .$$

Figure 6: Strength verification



Strength verification

The aim of strength verification is to dimension components reliably and appropriately for the material involved. Strength verification in accordance with Figure 6 is conducted in four stages:

1. Determination of the external load (forces and moments),
2. Calculation of the existing stress,
3. Choice of the material characteristic value,
4. Comparison of the existing stress with the material characteristic value.

If the external forces and moments are known, the stresses in the component can be calculated in accordance with Table 4. Because tensile, compressive and bending stresses (normal stresses) lie in one plane or act in the same direction, they can be cumulatively superimposed. Even the shear and torsional stresses can be added. However, if normal and shear stresses occur simultaneously in a component, a reduced stress must be created with a strength hypothesis because the material characteristic values were calculated from mono-axial tensile and vibration fatigue tests.

There are different hypotheses to suit the material behavior (ductile or brittle). For ductile materials (most metals) the deformation energy hypothesis is used. For the mono-axial stress state (bar) the reduced stress is calculated in accordance with Table 4.

Surface pressure

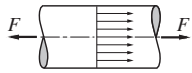
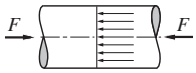
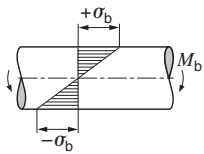
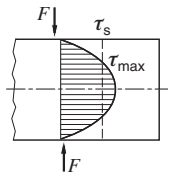
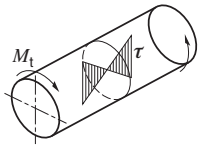
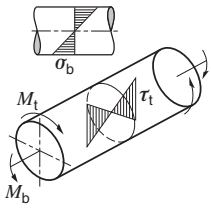
Surface pressure is a compressive stress. It is generated when a force F is transmitted from one solid body to another. For effective surface areas pressure p is equal to the ratio of force F to contact surface A.

Pressure on the face of a hole

When journal and bore or shaft and friction bearing are paired, the pressure on the face of a hole is usually used for calculation. Here the load is referred to the projected area A_{proj} (Figure 7):

$$p = \frac{F}{A_{proj}} = \frac{F}{dl} \leq p_{perm} \ .$$

Table 4: Basic equations of strength calculation

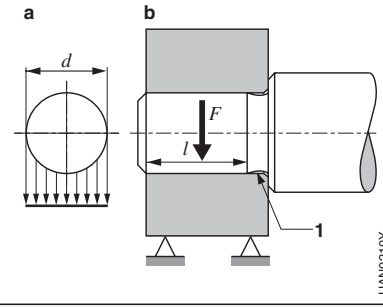
Loading	Stress
Tensile stress 	$\sigma_z = \frac{F}{A}$
Compressive stress  UAN0214Y	$\sigma_d = \frac{F}{A}$
Bending stress  UAN0215Y	$\sigma_b = \frac{M_b}{W_b}$ W_b from Table 5
Shear stress  UAN0216Y	$\tau_s = \frac{F}{A}$
Torsional stress  UAN0217Y	$\tau_t = \frac{M_t}{W_t}$ W_t from Table 5
Reduced stress  UAN0218Y	$\sigma_v = \sqrt{\sigma^2 + 3\tau^2}$

Hertzian stress

In reality, however, the maximum stress where the surfaces are curved is greater and can be calculated according to Hertz's theory. The maximum Hertzian stress is dependent on the deformation (flattening) of the touching surfaces (Figure 8). This in turn is dependent on the radii, on the modulus of elasticity E and

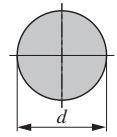
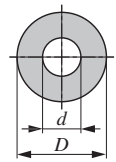
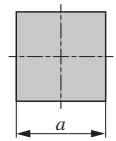
Figure 7: Pressure on the face of a hole

a) Projection,
b) Side view.
1 Undercut.
F Load,
d Journal diameter,
l Common contact length.



UAN0210Y

Table 5: Section moduli

Cross-section	Bending	Torsion
	$W_b = \frac{\pi}{32} d^3$	$W_t = \frac{\pi}{16} d^3$ UAN0219Y
	$W_b = \frac{\pi}{32}\left(\frac{D^4 - d^4}{D}\right)$	$W_t = \frac{\pi}{16}\left(\frac{D^4 - d^4}{D}\right)$ UAN0220Y
	$W_b = \frac{a^3}{6}$	$W_t = 0.208\, a^3$ UAN0221Y

on the transverse contraction ν. The equations for the cases "sphere – sphere" and "cylinder – cylinder" are set out in Table 6. For the special cases "sphere – plane" and "cylinder – plane" radius $r_2 \rightarrow \infty$ or $r = r_1$.

Figure 8: Hertzian stress
F Normal force,
r Radius,
p_{max} Maximum pressure.

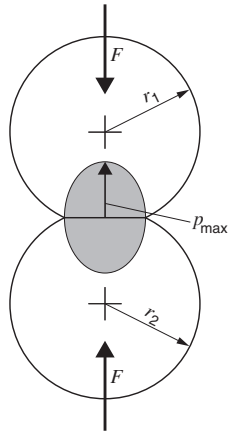


Table 6: Hertzian stress

Sphere – sphere (localized contact) $p_{max} = \frac{1}{\pi}\sqrt[3]{\frac{1.5FE^2}{r^2(1-\nu^2)^2}}$	Mean E modulus: $E = 2 \cdot \frac{E_1 E_2}{E_1 + E_2}$ For the radii: $\frac{1}{r} = \frac{1}{r_1} + \frac{1}{r_2} \rightarrow r = \frac{r_1 r_2}{r_1 + r_2}$
Cylinder – cylinder (line contact) $p_{max} = \sqrt{\frac{FE}{2\pi r l (1-\nu^2)}}$	l Contact length, cylinder ν Transverse contraction

Friction

Coulomb friction

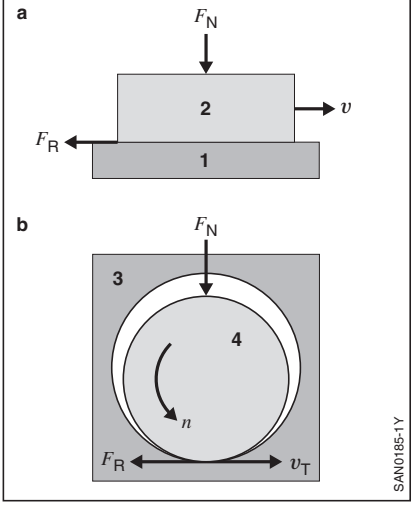
When touching bodies move relative to each other, friction acts as mechanical resistance acting in the opposite direction to the motion at velocity v (Figure 9). The force of resistance, known as frictional force F_R, is proportional to the normal force F_N. Static friction exists as long as the external force is less than the frictional force and the body remains at rest. When static friction is overcome, and the body is set in motion, frictional force is governed by Coulomb's law of sliding friction:

$$F_R = \mu F_N .$$

Coefficient of friction

The coefficient of friction μ always denotes a system property and not a material property. Coefficients of friction are, among other things, depending on material pairing (see Table 7), temperature,

Figure 9: Coulomb friction
a) Linear motion,
b) Rotary motion.
1 Base, 2 Sliding body,
3 Bearing shell, 4 Shaft.
F_N Normal force, F_R Friction force,
v Velocity, n Rotational speed,
v_T Tangential velocity at the point of contact.



surface condition, sliding velocity, the surrounding medium (e.g. water or CO_2, which can be adsorbed by the surface), and the intermediate material (lubricant). For this reason, coefficients of friction always fluctuate between limit values and may have to be calculated experimentally. Static friction is generally greater than sliding friction. In special cases, the friction coefficient can exceed 1, for example with very smooth surfaces where cohesion forces are predominant or with racing tires featuring an adhesion or suction effect.

Friction on the wedge
Allowing for friction, the insertion force is governed by (Figure 10):

$$F = 2\left(F_N \sin\frac{\gamma}{2} + F_R \cos\frac{\gamma}{2}\right)$$

The normal force is then

$$F_N = \frac{F}{2\left(\sin\frac{\gamma}{2} + \mu\cos\frac{\gamma}{2}\right)}.$$

Rope friction
Motions and moments can be transmitted with elastic, pliable ropes or belts (Figure 11). Sliding friction occurs in the event of relative motion between rope and pulley (e.g. belt brake or bollard with running rope). Static friction exists when rope and pulley are at rest relative to each other (e.g. belt drive, belt brake as holding brake, bollard with rope at rest). The coefficient of sliding friction μ or the static friction μ_H must be applied accordingly.

According to Euler's rope-friction formula, the force in the pulling strand (taut strand):

$$F_1 = F_2 e^{\mu\beta}.$$

The peripheral force F_U or frictional force F_R is calculated as

$$F_U = F_R = F_1 - F_2$$

and the transmittable friction torque as

$$M_R = F_R\, r.$$

Rolling friction
Rolling friction occurs when a ball, a caster or a wheel rolls on a track or roadway. Typical examples of this are the roller bearing, the pairing of wheel flange and rail for railroads or the pairing of tire and roadway for motor vehicles. Both rolling body and base are subjected to elastic deformation during rolling. This generates asymmetrical pressure (Figure 12). The equilibrium condition is used to calculate

Figure 10: Friction on the wedge
F_N Normal force,
F_R Frictional force,
γ Wedge angle.

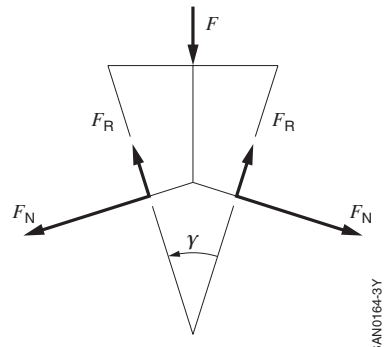
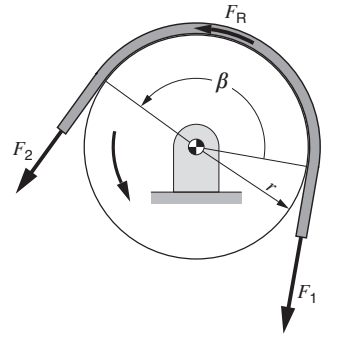


Figure 11: Rope friction
F_1 Force in taut strand,
F_2 Force in slack strand,
r Radius of pulley,
β Wrap angle.



frictional force F_R for the dry state (i.e. without lubrication):

$$F_R = \frac{x}{R} F_N = \mu_R F_N .$$

Ratio x/R can be calculated as frictional resistance or coefficient of rolling friction μ_R. From this it can be seen that large rolling bodies roll more easily than small rolling bodies. Hard surfaces (roller bearing and railroad) give rise to small deformations and therefore result in very low coefficients of friction; soft surfaces, as in the tire/roadway pairing, on the other hand result in higher coefficients of friction. While frictional resistances of μ_R = 0.0015 can be achieved for ball bearings, the coefficients of friction of a car tire on asphalt are μ_R = 0.015.

References
[1] A. Böge: Technische Mechanik. 30th Ed., Verlag Springer Vieweg, 2013.
[2] R.C. Hibbeler: Technische Mechanik 1, 2 und 3, Pearson Studium.
[3] K.-H. Grote, J. Feldhusen: Dubbel – Taschenbuch für den Maschinenbau. 23rd Ed.; Springer-Verlag 2011.

Figure 12: Rolling friction
a) Asymmetrical pressure distribution,
b) Resultant forces for calculation model.
F Load, F_N Normal force, F_R Frictional force,
R Radius of rolling body,
x Lever arm for moment equilibrium.

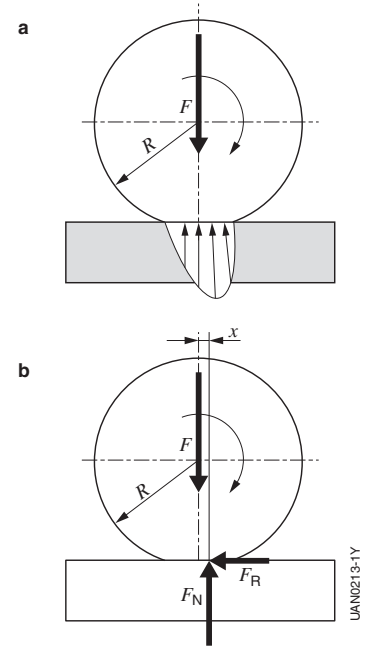


Table 7: Reference values for coefficients of static and sliding friction

Material pair	Coefficient of static friction μ_H		Coefficient of sliding friction μ	
	Dry	Lubricated	Dry	Lubricated
Steel – steel	0.15 – 0.20	0.10	0.10 – 0.15	0.05 – 0.10
Steel – cast iron	0.18 – 0.25	0.10	0.15 – 0.20	0.10
Steel – sintered bronze	0.20 – 0.40	0.08 – 0.13	0.18 – 0.30	0.06 – 0.09
Steel – brake pad	– [1]	– [1]	0.50 – 0.60	0.20 – 0.50
Steel – polyamide	0.60	0.20	0.32 – 0.45	0.10
Steel – ice	0.027	– [1]	0.014	– [1]
Cast iron – cast iron	0.18 – 0.25	0.10	0.15 – 0.20	0.10
Aluminum – aluminum	0.50 – 1.00	– [1]	0.50 – 1.00	– [1]
Wood – wood	0.40 – 0.60	0.20	0.20 – 0.40	0.10
Wood – metal	0.60 – 0.70	0.10	0.40 – 0.50	0.10
Leather – metal	0.50 – 0.60	0.20 – 0.25	0.20 – 0.25	0.12
Car tire on dry asphalt	0.50 – 0.60	– [1]	– [1]	– [1]
Car tire on wet asphalt	0.20 – 0.30	– [1]	– [1]	– [1]
Car tire on icy asphalt	0.10	– [1]	– [1]	– [1]

[1] No practical use.

Fluid mechanics

Hydrostatics

Density and pressure

Although fluids are compressible to a lesser extent, they can be viewed as being incompressible for most problems. In addition, since density is only slightly dependent on temperature, it can be taken as constant for many applications.

Pressure $p = dF/dA$ is non-directional in fluids which are at rest. If the pressure component produced by the difference in height (geodetic pressure) is negligible, the hydrostatic pressure is uniformly high everywhere (e.g. in a hydrostatic press).

Table 1: Symbols and units

Quantity		Unit
A	Cross-sectional area	m^2
A_B	Area of base	m^2
A_S	Area of side	m^2
F	Force	N
F_A	Buoyancy force	N
F_B	Force acting on bottom	N
F_G	Weight	N
F_S	Lateral force	N
F_W	Volume of displaced fluid	N
L	Length in flow direction	m
Q	Volumetric flow	m^3/s
R_e	Reynolds number	–
V_F	Volume of displaced fluid	m^3
c_W	Drag coefficient	–
d	Diameter	m
g	Acceleration of free fall ($g \approx 9{,}81\ m/s^2$)	m/s^2
h	Depth of fluid	m
m	Mass	kg
m_F	Mass of displaced fluid	kg
$\dot{m}$	Mass flow	kg/s
p	Pressure	$Pa = N/m^2$
t	Thickness	m
w	Flow velocity	m/s
α	Contraction coefficient	–
η	Dynamic viscosity	$Pa \cdot s = Ns/m^2$
μ	Discharge coefficient	–
ν	Kinematic viscosity	m^2/s
ρ	Density	kg/m^3
φ	Velocity coefficient	–
τ	Shear stress	N/m^2

Fluid at rest in an open vessel

In the case of open vessels, the pressure in the fluid is only dependent on the depth of the fluid (Figure 1).Closed vessels with pressure compensation, such as fuel tanks and brake-fluid reservoirs, can also be considered as open vessels. The following applies:

Pressure: $p(h) = \rho g h$
Force acting on bottom: $F_B = A_B \rho g h$
Force acting on sides: $F_S = 0.5 A_S \rho g h$

Hydrostatic press

By way of example, power amplification in vehicle brakes and hydraulic power-assisted steering systems functions according to the hydrostatic-press principle (Figure 2). The following relationships apply to the pressure p and the piston forces F:

Figure 1: Pressure distribution in fluid at rest
p Pressure,
h Depth of fluid.

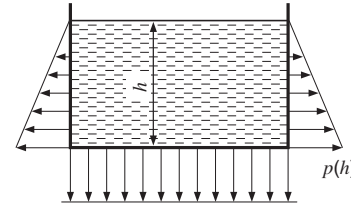
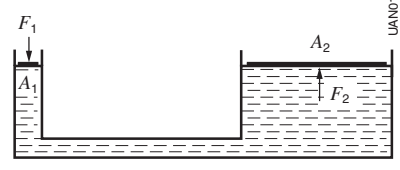


Figure 2: Hydrostatic press
F Force,
A Cross-sectional area.



Pressure: $p = \frac{F_1}{A_1} = \frac{F_2}{A_2}$.

Piston forces: $F_1 = pA_1 = F_2\frac{A_1}{A_2}$,

$$F_2 = pA_2 = F_1\frac{A_2}{A_1}.$$

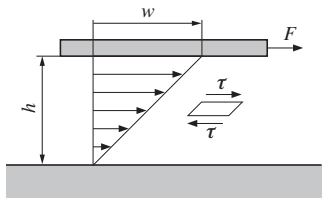
Buoyancy
Buoyancy is a force acting against gravity and acts on the center of gravity of the volume of the displaced fluid. It corresponds to the weight of the fluid displaced by the submerged body:

$F_A = m_F g = V_F \rho g$.

A body will float if $F_A = F_G$.

The amount of fuel available can be easily and reliably measured by analog sensors (floats) in the fuel tank with the aid of buoyancy.

Figure 3: Shear stresses in fluids
τ Shear stress,
w Flow velocity,
h Height,
F Force.



Flow mechanics

Basic principles
An ideal fluid (generic term for gases and liquids) is incompressible and frictionless. This means that no shear stresses occur in the fluid, and the pressure on a fluid element is uniform in all directions. In actual fact, however, a resistance must be overcome in fluids if deformations occur that are caused by displacement of fluid elements (Figure 3). The resulting shear stress complies with Newton:

$$\tau = \frac{F}{A} = \eta\,\frac{w}{h}.$$

The proportionality factor η is called the dynamic viscosity and is greatly dependent on temperature. In practice, however, the kinematic viscosity

$$\nu = \frac{\eta}{\rho}$$

is often used, as it can be measured very easily with a capillary viscometer.

Flows without turbulence, in which the individual fluid layers move separately in parallel and which are predominantly determined by viscosity, are known as laminar flows. If the flow velocity exceeds a limit value, adjacent layers start to swirl, resulting in a turbulent flow. In addition to the flow velocity, the transition point between laminar and turbulent flow is also dependent on the Reynolds number

$$R_e = \frac{\rho L w}{\eta} = \frac{L w}{\nu}$$

In the case of a flow within a pipe, the pipe diameter is used for L. Flow in a pipe becomes unstable or turbulent at $R_e > 2300$.

Since, in the case of gases with low flow velocities (up to 0.5 times the velocity of sound), compression is negligible in many flow processes, they are also governed by the laws of incompressible fluids.

Basic equations of flow mechanics

The most important basic equations of flow mechanics are the continuity equation and the Bernoulli equation. They describe the conservation of mass and energy in flowing fluids.

Continuity equation

In a steady state, mass conservation requires that in a flow the mass flow rate be of equal magnitude in each cross-section (Figure 4):

$$\dot{m} = \rho A_1 w_1 = \rho A_2 w_2 = \text{const.}$$

In the case of incompressible fluids (ρ = konst.), the volumetric flow must also be constant:

$$Q = A_1 w_1 = A_2 w_2 = \text{konst.}$$

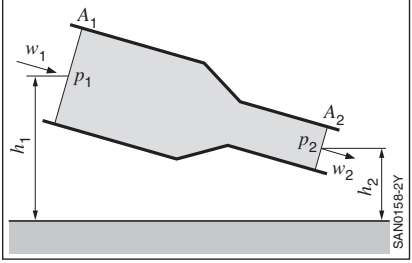
Bernoulli equation

From the continuity equation, it follows that an acceleration takes place between A_1 and A_2. This results in an increase in kinetic energy, which must be effected by a pressure drop, where $p_1 > p_2$ (Figure 4). According to the law of conservation of energy, the sum of the static pressure p, kinetic pressure, and geodetic pressure is constant in a flowing fluid. Ignoring friction losses, the following applies accordingly to the flowing fluid in a non-horizontal pipe:

$$p_1 + \frac{1}{2}\rho w_1^2 + \rho g h_1 = p_2 + \frac{1}{2}\rho w_2^2 + \rho g h_2 .$$

Figure 4: Continuity and Bernoulli equations

A Cross-sectional area, h Height, p Pressure, w Flow velocity.



Discharge from a pressure vessel

Under the precondition that the cross-sectional area of the discharge end is very much smaller than that of the vessel (Figure 5), the velocity w_1 is negligible according to the continuity equation. As derived from Bernoulli equation, the discharge velocity is governed by the following:

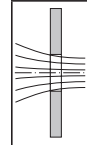
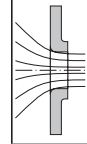
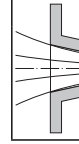
$$w_2 = \varphi\sqrt{\frac{2}{\rho}(p_1 - p_2) + 2gh} .$$

The velocity coefficient φ takes into account the losses that occur. The jet constriction must also be taken into account for the volumetric flow or the discharge volume; this constriction is dependent on the contraction coefficient α. The following then applies to the discharge volumetric flow:

$$Q = \alpha\varphi A_2\sqrt{\frac{2}{\rho}(p_1 - p_2) + 2gh} .$$

Velocity coefficient and contraction coefficient are often expressed together as the discharge coefficient $\mu = \alpha\varphi$ (Table 2).

Table 2: Discharge openings

Orifice shape	Velocity coefficient φ	Contraction coefficient α				Discharge coefficient μ
	0.97	0.61…0.64				0.59 … 0.62
	0.97 … 0.99	1.0				0.97 … 0.99
 SAN0159-1Y		$(d_2/d_1)^2$				
		0.4	0.6	0.8	1.0	
	0.95 … 0.97	0.87	0.90	0.94	1.0	0.82 … 0.97

Resistance of bodies submerged in a fluid flow

A pressure differential occurs across a body submerged in a fluid flow (e.g. the vehicle body), resulting in a resistance force

$$F_W = \frac{1}{2} c_W A \rho w^2 .$$

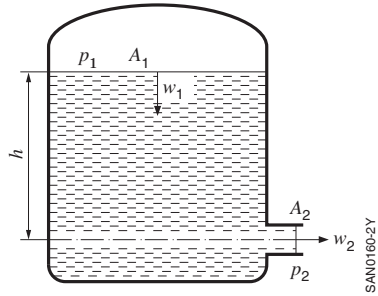
Here, A is the cross-sectional area of the body on which the fluid flow is acting and c_d an undefined coefficient of resistance, which is dependent on the shape of the body submerged in the fluid flow.

As it is extremely complex to calculate exactly the resistance to flow even for simple bodies, resistance to flow is usually determined experimentally. If the dimensions are large, the measurements are taken on downscaled models. As well as geometrical similarity, the forms of energy that occur (kinetic energy, frictional work) in the original fluid flow and in the model flow must be proportional. This proportion is denoted by the Reynolds number R_e.

Basically: Two flows are similar in fluid-dynamic terms if their Reynolds numbers R_e are identical. Because even complex geometries are composed of simple basic bodies, streamlined surfaces can already be derived during the modeling phase as per Table 3.

Figure 5: Discharge from a pressure vessel

A Cross-sectional area, h Height, p Pressure, w Flow velocity.



References

[1] A. Böge; W. Böge: Technische Mechanik. 31st Ed., Verlag Springer Vieweg, 2015.

[2] W. Bohl; W. Elmendorf: Technische Strömungslehre. 15th Ed., Vogel-Verlag, 2014.

Table 3: Drag coefficients c_W

Body shape: *L Length, t Thickness,* *R_e Reynolds number.*	c_W
Circular plate	1.11
Open dish	1.33
Sphere R_e < 200,000 R_e > 250,000	 0.47 0.20
Narrow rotational body L/t = 6	0.05
Long cylinder R_e < 200,000 R_e > 450,000	 1.0 0.35
Long plate L/t = 30 $R_e \approx 500{,}000$ $R_e \approx 200{,}000$	 0.78 0.66
Long wing L/t = 18 $R_e \approx 10^6$ L/t = 8 $R_e \approx 10^6$ L/t = 5 $R_e \approx 10^6$ L/t = 2 $R_e \approx 2 \cdot 10^5$	 0.2 0.1 0.08 0.2

SAN0161-2Y

Vibrations and oscillations

Terms

(Symbols and units in Table 1, see also DIN 1311, [1] and [2]).

Vibration or oscillation
A vibration or oscillation is the change in a physical quantity which repeats at more or less regular time intervals and whose direction changes with similar regularity (Figure 1).

Period
The period T is the time taken for one complete cycle of a single oscillation.

Amplitude
The amplitude $\hat{y}$ is the maximum instantaneous value (peak value) of a sinusoidally oscillating physical quantity.

Frequency
The frequency f is the number of oscillations in one second, the reciprocal value of the period of oscillation T.

Angular frequency
The angular frequency ω is 2π times the frequency f.

Particle velocity
Particle velocity v is the instantaneous value of the alternating velocity of a vibrating particle in its direction of vibration. It must not be confused with the velocity of propagation of a traveling wave (e.g. the velocity of sound).

Figure 1: Sinusoidal oscillation
(Quantities, see Table 1)

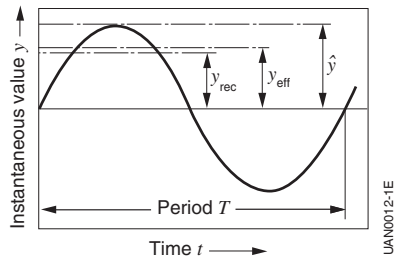


Table 1: Symbols and units

Quantity		Unit
a	Storage coefficient	
b	Damping coefficient	
c	Storage coefficient	
c	Spring constant	N/m
c_α	Torsional rigidity	N m/rad
C	Capacitance	F
f	Frequency	Hz
f_g	Resonant frequency	Hz
Δf	Half-value width	Hz
F	Force	N
F_Q	Excitation function	
I	Current	A
J	Moment of inertia	$kg \cdot m^2$
L	Inductance	H
m	Mass	kg
M	Torque	N m
n	Rotational speed	1/min
Q	Charging	C
Q	Resonance sharpness	
r	Damping factor	N s/m
r_α	Rotational damping coefficient	N s · m
R	Ohmic resistance	Ω
t	Time	s
T	Period	s
U	Voltage	V
v	Particle velocity	m/s
x	Travel/displacement	m
y	Instantaneous value	
$\hat{y}$	Amplitude	
$\dot{y}$ ($\ddot{y}$)	Single (double) derivative with respect to time	
y_{rec}	Rectification value	
y_{eff}	Effective value	
α	Angle	rad
δ	Decay coefficient	1/s
Λ	Logarithmic decrement	
ω	Angular velocity	rad/s
ω	Angular frequency	1/s
Ω	Excitation-circuit frequency	1/s
D	Damping ratio	
D_{opt}	Optimum damping ratio	
	Subscripts:	
0	Undamped	
d	Damped	
T	Absorber	
U	Base support	
G	Machine	

Fourier series
Every periodic function, which is piecewise monotonic and smooth, can be expressed as the sum of sinusoidal harmonic components.

Beats
Beats occur when two sinusoidal oscillations, whose frequencies do not differ greatly, are superposed ($f_1 \approx f_2$). They are periodic. Their basic frequency is the difference between the frequencies of the superposed sinusoidal oscillations.

Natural frequency
The natural frequency is that frequency f at which an oscillating system can oscillate freely after being excited once (natural oscillation). It is dependent only on the properties of the oscillating system.

Damping
Damping is a measure of the energy losses in an oscillatory system when one form of energy is converted into another. The consequence is a decay of the oscillation (Figure 2).

Damping ratio
The damping ratio D is the measure for the degree of damping.

Logarithmic decrement
The logarithmic decrement Λ is the natural logarithm of the relationship between two extreme values of a damped natural oscillation which are separated by one period.

Forced oscillations
Forced oscillations arise under the influence of an external physical force (excitation), which does not change the properties of the oscillator. The frequency of forced oscillations is determined by the frequency of the excitation.

Transfer function
The transfer function is the quotient of amplitude of the observed state or output variable and the amplitude of excitation, plotted against the excitation frequency f or the excitation-circuit frequency ω.

Resonance
A resonance occurs when the transfer function attains its maximum value as the excitation frequency approaches the natural frequency.

Resonant frequency
The resonant frequency is the excitation frequency at which the oscillator state variable attains its maximum value. Disregarding the damping, the resonant frequency is equal to the natural frequency.

Half-value width
The half-value width is the difference between the frequencies at which the level of the observed variable has dropped to $1/\sqrt{2} \approx 0.707$ of the maximum value.

Resonance sharpness
The resonance sharpness Q (or quality factor) is the maximum value of the transfer function.

Coupling
If two oscillatory systems are coupled together – mechanically by mass or elasticity, electrically by inductance or capacitance – a periodic exchange of energy takes place between the systems.

Wave
A wave is a spatial and temporal change of state of a continuum, which can be expressed as a unidirectional transfer of location of a certain state over a period of time. The matter that may be present in the space is not necessarily also transported.

There are transversal waves (e.g. waves in rope and water) and longitudinal waves (e.g. sound waves in air).

Interference
The principle of undisturbed superposition of waves is called interference. At every point in space, the instantaneous value of the resulting wave is equal to the sum of the instantaneous values of the individual waves.

Plane wave
A plane wave is a wave in which the surfaces of the same phase (e.g. maxima or wave fronts) form a plane, i.e. the wave propagates linearly. The wave fronts are vertical to the direction of propagation.

Standing waves
Standing waves occur as a result of interference between two waves of equal frequency, wavelength, and amplitude traveling in opposite directions. In contrast to a propagating wave, the amplitude of the standing wave is constant at every point; nodes (zero amplitude) and antinodes (maximum amplitude) occur. Standing waves occur, for example, when a plane wave is reflected on a plane wall which is vertical to the direction of wave propagation.

Rectification value
The rectification value y_{rec} is the arithmetic mean value, linear in time, of the values of a periodic signal:

$$y_{rec} = \frac{1}{T}\int_0^T |y|\, dt.$$

For a sine curve:

$$y_{rec} = \frac{2\hat{y}}{\pi} \approx 0.637\,\hat{y}.$$

Effective value
The effective value y_{eff} is the time virtual value of a periodic signal. It is also known as the RMS value (root mean square):

$$y_{eff} = \sqrt{\frac{1}{T}\int_0^T y^2\, dt}.$$

For a sine curve:

$$y_{eff} = \frac{\hat{y}}{\sqrt{2}} \approx 0.707\,\hat{y}.$$

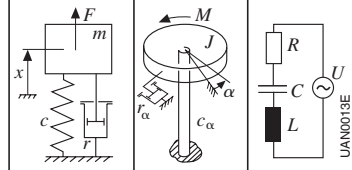
Harmonic factor
The harmonic factor is the ratio of y_{eff} to y_{rec}. For a sine curve the harmonic factor is $y_{eff}/y_{rec} \approx 1.111$.

Peak factor
For a sine curve the peak factor is $\hat{y}/y_{eff} = \sqrt{2} \approx 1.414$.

Equations

The following equations apply to simple oscillators (Table 2) if the general quantity designations in the formulas are replaced by the relevant physical quantities.

Table 2: Simple oscillatory systems

	Mechanical		Electrical
	Translational	Rotational	
			
Des.	Physical quantity		
y	x	α	Q
$\dot{y}$	$\dot{x} = v$	$\dot{\alpha} = \omega$	$\dot{Q} = I$
$\ddot{y}$	$\ddot{x} = \dot{v}$	$\ddot{\alpha} = \dot{\omega}$	$\ddot{Q} = \dot{I}$
F_Q	F	M	U
a	m	J	L
b	r	r_α	R
c	c	c_α	$1/C$

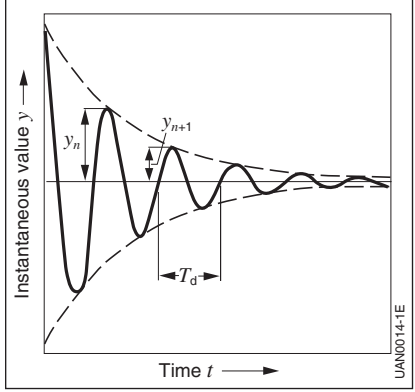
Differential equation

$$a\ddot{y} + b\dot{y} + cy = F_Q(t) = \hat{F}_Q \sin \Omega t,$$

Period $T = 1/f$,
Angular frequency $\omega = 2\pi f$.

Sinusoidal oscillation: $y = \hat{y} \sin \omega t$.

Figure 2: Free oscillation and damping for $0 < D < 1$



Free oscillations ($F_Q = 0$)
Logarithmic decrement (Figure 2):

$$\Lambda = \ln\left(\frac{y_n}{y_{n+1}}\right) = \frac{\pi b}{\sqrt{ca - \frac{b^2}{4}}},$$

Decay coefficient $\delta = \frac{b}{2a}$,

Damping ratio $D = \frac{\delta}{\omega_0} = \frac{b}{2\sqrt{ca}}$,

$$D = \frac{\Lambda}{\sqrt{\Lambda^2 + 4\pi^2}} \approx \frac{\Lambda}{2\pi} \text{ (low level of damping).}$$

Angular frequency of undamped oscillation ($D = 0$): $\omega_0 = \sqrt{c/a}$.

Angular frequency of damped oscillation ($0 < D < 1$): $\omega_d = \omega_0\sqrt{1-D^2}$.

For $D \geq 1$ no oscillations, but creepage.

Forced oscillations
Quantity of transfer function:

$$\frac{\hat{y}}{\hat{F}_Q} = \frac{1}{\sqrt{(c - a\Omega^2)^2 + (b\,\Omega)^2}}$$

$$= \frac{1}{c\sqrt{\left(1 - \left(\frac{\Omega}{\omega_0}\right)^2\right)^2 + \left(\frac{2D\Omega}{\omega_0}\right)^2}},$$

Resonant frequency $f_g = f_0\sqrt{1-2D^2} < f_0$,

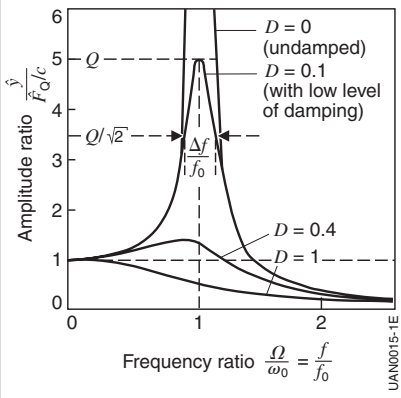
Resonance sharpness $Q = 1/\left(2D\sqrt{1-D^2}\right)$,

Resonant frequency $f_g \approx f_0$ (for $D \leq 0.1$),

Resonance sharpness $Q \approx \frac{1}{(2D)}$ (for $D \leq 0.1$),

Half-value width $\Delta f = 2Df_0 = \frac{f_0}{Q}$.

Figure 3: Standardized transfer function



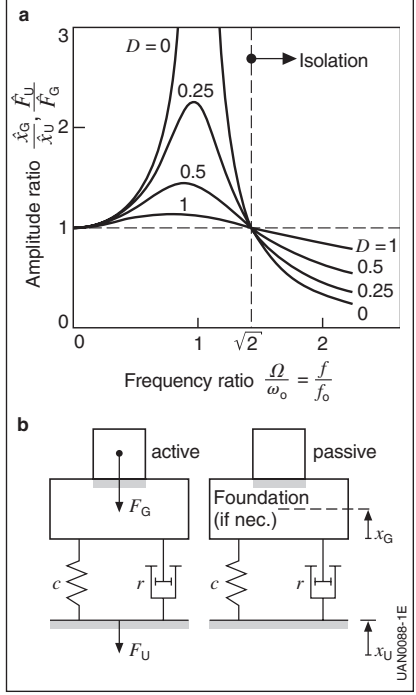
Vibration reduction

Vibration damping
If damping can only be carried out between the machine and a quiescent point, damping must be at a high level (Figure 3).

Vibration isolation
Active vibration isolation
The machines to be isolated are to be mounted so that the dynamic forces transmitted to the base support are small.

One measure to be taken: The bearing point should be set below resonance so that the natural frequency lies below the lowest excitation frequency. Damping impedes insulation. Low values can result in excessively high vibrations during run-up when the resonant range is passed through (Figure 4).

Figure 4: Vibration isolation
a) Transfer function,
b) Low setting.

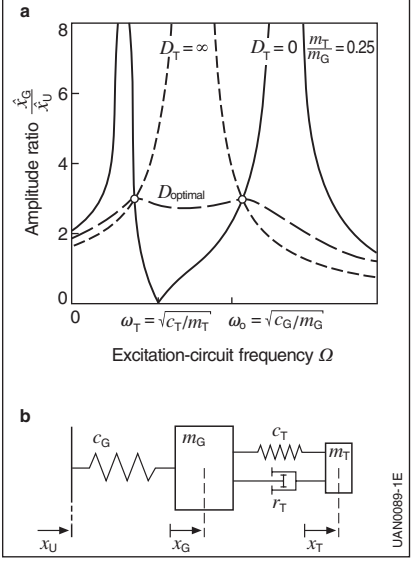


Passive vibration isolation
The machines to be isolated are to be mounted so that vibrations and shocks reaching the base support are only transmitted to the machines to a minor extent.

The measures to be taken are the same as those for active isolation. In many cases, flexible suspension or extreme damping is not practicable. To prevent the occurrence of resonance, the machine attachment should be so rigid that the natural frequency is far enough in excess of the highest excitation frequency which can occur (Figure 4).

Vibration absorption
Vibration absorber
with fixed natural frequency
By tuning the natural frequency f_T of an absorption mass with a flexible, loss-free coupling to the excitation frequency, vibrations acting on the machine are completely absorbed (Figure 5). Only the absorption mass still vibrates. The effectiveness of the absorption decreases as the excitation frequency changes. Damping prevents complete absorption. However, appropriate tuning of the absorber frequency and an optimum damping ratio produce broadband vibration reduction, which remains effective when the excitation frequency changes.

Figure 5: Vibration absorption
a) Transfer function of machine,
b) Schematic.



Vibration absorber
with varying natural frequency
Rotational oscillations with excitation frequencies proportional to the rotational speed (e.g. orders of balancing in IC engines) can be absorbed by absorbers with natural frequencies proportional to the rotational speed (pendulum in the centrifugal-force field). The vibration absorption is effective at all rotational speeds. Vibration absorption is also possible for oscillators with several degrees of freedom and interrelationships, as well as by the use of several absorption masses.

Modal analysis

The dynamic behavior (natural-oscillation characteristics) of oscillating systems is determined by means of modal analysis. It is used among other things in design to optimize structures with regard to the oscillatory characteristics and to identify problem areas, and in acoustics to analyze structure-borne noise ([3]).

The oscillating structure, which as a continuum has infinitely many degrees of freedom, is replaced in a clearly defined manner by a finite number of single-mass oscillators. The modal model of the structure created in this way is described by the modal parameters
– natural-oscillation shapes (also eigenvector or mode),
– natural frequencies (also eigenvalues),
– and the associated modal damping values.

A time-invariant and linear-elastic structure is an essential precondition for model creation. Every oscillation of the structure can be represented from the eigenvectors and eigenvalues. It is, however, only observed at a limited number of points in the possible oscillation directions (degrees of freedom) and at defined frequency intervals.

The substructure-coupling process collates modal models of various structures into an overall model.

Numerical modal analysis
The geometry, material data, and marginal conditions must be known. The basis for numerical modal analysis is a multibody-system or finite-element model of the structure. Eigenvalues and eigenvectors can be calculated from this by solving an eigenvalue problem.

Numerical modal analysis manages without a prototype of the structure and can already be used at an early stage of development. However, it is often the case that precise knowledge concerning the structure's fundamental properties (damping, marginal conditions) are lacking, which means that the modal model can sometimes be inaccurate. As well as this, the error is unidentified. One remedy is to adjust the model with the results of an experimental modal analysis.

Experimental modal analysis
A prototype of the structure is required for experimental modal analysis. Analysis is based on measurements of transfer functions. To this end, either the structure is excited at one point in the frequency range of interest and the oscillation responses measured at several points, or it is excited at many points in succession, where the oscillation responses are always measured at the same point. The modal model is derived from the matrix of the transfer function (it describes the response model). An impulse hammer or an electrodynamic or hydraulic "shaker" is used as the means of excitation. The response is measured with acceleration sensors or a laser vibrometer.

Experimental modal analysis can also be used to validate numerical analysis. Simulation calculations can then be carried out on the validated numerical model. In the response calculation, the response of the structure to a defined excitation is calculated, an excitation which corresponds, for example, to test-bay conditions.

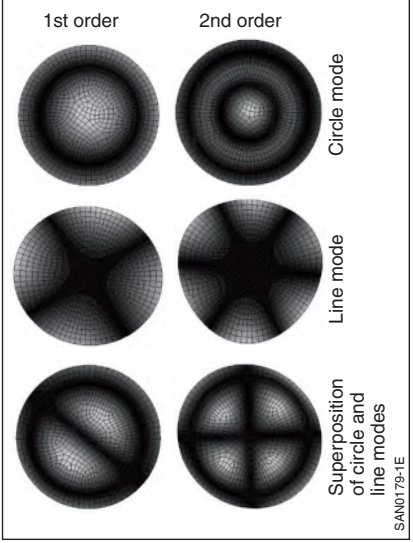
By means of structure modifications (changes in mass, damping or stiffness), the vibrational behavior can be optimized to the level required by operating conditions. When the modal models produced by both processes are compared with each other, the modal model resulting from an analytical modal analysis is more detailed than that from an experimental modal analysis, due to the greater number of degrees of freedom in the analytical process. This applies in particular to simulation calculations based on the model.

The natural-oscillation shapes resulting from a modal analysis can be shown in graphic form or animated (examples in Figure 6). The different gray stages describe the excursion vertically to the plane of projection. The partial distortion of the disks results from these excursions.

References

[1] DIN 1311: (Mechanical) vibrations, oscillation and vibration systems – Part 1: Basic concepts, survey.
[2] P. Hagedorn, D. Hochlenert: Technische Schwingungslehre. 2nd Ed., Europa Lehrmittel, 2014.
[3] H.G. Natke: Einführung in Theorie und Praxis der Zeitreihen- und Modalanalyse. 3rd Ed., Vieweg+Teubner Verlag, 2013.

Figure 6: Eigenmode of round disks



Acoustics

Acoustics (Greek “to hear”) is a branch of the general science of vibration and oscillation concerned with the vibration of air particles audible to humans. This process of vibration gives rise to pressure differences in air which the human ear can detect. Vibrations are generated on an automobile by a variety of excitations which partly excite the airborne noise directly or can propagate via the vehicle structure and from there to the airborne noise.

Table 1: Quantities and units
(see also DIN EN ISO 80000-8 [2])

Quantity		SI unit
c	Velocity of sound	m/s
f	Frequency	Hz
I	Sound intensity	W/m^2
L_I	Sound intensity level	dB
L_Aeq	Equivalent continuous sound level, A-weighted	dB (A)
L_den	Noise indices	dB (A)
L_pA	Sound pressure level, A-weighted	dB (A)
L_r	Rating sound level	dB (A)
L_WA	Sound power level, A-weighted	dB (A)
L_S	Loudness level	phon
S	Loudness	sone
P	Sound power	W
p	Sound pressure	Pa
S	Surface area	m^2
v	Particle velocity	m/s
Z	Acoustic impedance	Pa · s/m
α	Sound absorption coefficient	1
λ	Wavelength	m
ρ	Density	kg/m^3
ω	Angular frequency (= 2 πf)	1/s
SEL	Sound exposure level	dB (A)

General terminology

(Symbols and units in Table 1, see also DIN 1320 [1])

Sound
Mechanical vibrations and waves in an elastic medium in the audible frequency band (16 to 20,000 Hz) are called sound.

Ultrasound
High-frequency vibrations above the frequency range of human hearing are called ultrasound.

Infrasound
Infrasound is the range of very low frequencies which can no longer be detected by the human hear, but can still be felt by the body.

Sound pressure
Sound pressure p is the alternating pressure generated in a medium by the vibration of sound.

Particle velocity
Particle velocity v is the alternating velocity of a vibrating particle. In a free sound field:

$$v = \frac{p}{Z}\,.$$

Specific acoustic impedance
Specific acoustic impedance Z denotes the wave impedance of a medium and is a measure of the ability of a medium to transmit sound waves.

In a free field the relationship between sound pressure, particle velocity and acoustic impedance is:

$$Z = \frac{p}{v} = \rho\, c\,.$$

Figuratively speaking, the formula describes the coupling of a vibrating particle to its neighboring particles. The coupling is much less pronounced in gases than in solid media; the particles can vibrate more freely.

Hearing dynamics
Hearing dynamics is the ability of the human ear to perceive noises from the threshold of audibility up to the pain threshold. It corresponds to pressure differences from the threshold of audibility, which is generally defined as 20 µPa, up to the pain threshold of approx. 200 pascal. This results in very high hearing dynamics of around 1:10,000,000.

Decibel dB
The decibel is not a unit, but the logarithmic ratio of a measured quantity to a reference value. The use of decibels makes it easier to handle the large dynamic range.

The sound pressure level is given in dB referred to the reference pressure of the threshold of audibility (p_{ref} = 20 µPa):

$$L_p = 20 \log_{10}\left(\frac{p}{p_{ref}}\right).$$

Frequency
Frequency f specifies the number of vibrations per second. The unit is the hertz (Hz) and is defined as follows:

1 Hz = 1/s.

Velocity of sound
Velocity of sound c is the velocity of propagation of a sound wave in a medium (Table 2).

Table 2: Sound velocities and wavelengths in different materials

Material/medium	Sound velocity c in m/s	Wavelength λ in m at 1,000 Hz
Air, 20 °C, 1,014 hPa	343	0.343
Water, 10 °C	1,440	1.44
Rubber (acc. to hardness)	60–1,500	0.06–1.5
Aluminum (rod)	5,100	5.1
Steel (rod)	5,000	5.0

Wavelength
The distance between two wave crests is called the wavelength. It is linked to the frequency:

$$\lambda = \frac{c}{f} = \frac{2\pi c}{\omega}.$$

Propagation of sound
In general, sound propagates spherically from its source. In a free sound field, the sound pressure level decreases by 6 dB each time the distance from the sound source is doubled (20log2). Reflecting objects influence the sound field, and the rate at which the sound level is reduced as a function of the distance from the sound source is lower. Moving vehicles have a spherical propagation of sound, but are considered to be line sources. The damping of sound each time the distance is doubled is considerably less and ranges, depending on the vehicle and the computing model, between 3 dB and 4.5 dB each time the distance is doubled.

Sound power
Sound power P is the sound energy output by a sound source per unit of time. It is not dependent on space and distance.

Sound intensity
Sound intensity I is the sound power through a plane vertical to the direction of propagation:

$$I = \frac{P}{S}.$$

In a sound field:

$$I = \frac{p^2}{\rho} c = v^2 \rho c.$$

Doppler effect
For moving sound sources: As the sound source approaches the observer, the perceived pitch is higher than the actual pitch; as distance increases, the perceived pitch falls.

Sound spectrum
Every noise is a mixture of components with different frequencies and levels. The sound-pressure level components can be broken down by frequencies with a frequency analysis. These spectra are distinguished from each other, depending on the type of frequency resolution.

Octave band spectrum
The sound pressure levels are determined and represented in terms of octave bandwidth. In the case of the octave, the fundamental frequencies behave in a ratio (f_1, f_2) of 1:2. The mean frequency of the octave is:

$$f_m = \sqrt{f_1 f_2}\ .$$

Third-octave band spectrum
The sound pressure levels are determined and represented in terms of third-octave bandwidth. In the case of the third, the fundamental frequencies behave in a ratio of $1:2^{1/3}$. The bandwidth referred to the center frequency is relatively constant, as in the case of the octave band spectrum.

The frequencies are defined in DIN EN 61260 [3]. Frequency analyses in the octave and third-octave ranges produce bar charts with the frequency on the x-axis and the associated levels on the y-axis.

Narrow-band spectrum
Unlike the above spectra, it is possible to use Fourier analysis [4] to analyze frequency components with constant frequency bandwidths. This results in very many finer frequency resolutions compared with the above spectra. These are shown as line graphs, which strictly speaking are not correct, but enable different analyses to be compared.

Such narrow-band analyses are also frequently linked with a further piece of information such as, for example, the rotational speed of an engine so that a frequency analysis is available for each rotational speed (for example when the engine is revved up). Such analyses are shown in three-dimensional form, either as “waterfall charts” or today frequently in color spectra, with the x-axis showing the rotational speed, the y-axis the frequency, and the z-axis in its aspect the color coding.

Sound insulation
Sound insulation is achieved by interposing a reflecting (insulating) wall (e.g. a building wall) between the sound source and the impact location. This reduces the effect of the sound source.

Sound damping, sound absorption
In the case of sound damping and sound absorption, sound energy can penetrate into the medium. Here, energy is converted into heat when reflected on peripheries, but also during propagation in the medium.

Modern noise baffles work on the principle of sound damping, i.e. sound is reflected. But parts of the sound energy are also absorbed by the baffles.

Sound absorption coefficient
The sound absorption coefficient α is the ratio of non-reflected to incident sound energy. With total reflection, $\alpha = 0$; with total absorption, $\alpha = 1$.

Low-noise design
A low-noise design is a design which is structurally optimized in accordance with acoustic considerations in order to minimize the inherent sound radiation and the propagation of sound within the structure. Simulation techniques for calculating and optimizing acoustic properties are frequently used.

Noise reduction
Noise reduction involves reducing the noise emissions of a complete system, primarily by means of low-noise designs and, secondly, through the reduction of sound propagation through the use of insulating, damping and absorbing materials.

Measured quantities for noise emissions

Sound-field quantities are usually given as effective values. Because humans perceive frequencies with varying loudness, weighting filters are used to adapt the measured levels in each frequency range to the human hearing properties. The most common weightings are A for sound as encountered, for example, in the automobile sector, and C for clearly louder noises such as, for example, in the aviation field. The corresponding weighting is appended to the symbol, such as, for example, dB(A).

Sound power level
The sound power of a sound source is described by the sound power level L_w. It is equal to ten times the logarithm to the base 10 of the ratio of the calculated sound power to the reference sound power $P_0 = 10^{-12}$ W:

$$L_w = 10 \log\left(\frac{P}{P_0}\right),$$

where log here and also in the following means the logarithm to the base 10.

Sound power cannot be measured directly. It is calculated based on quantities of the sound field which surrounds the source. The sound pressure levels L_p or the sound intensity levels L_I are usually used for the calculation.

Sound pressure level
The sound pressure level L_p is ten times the logarithm to the base 10 of the ratio of the square of the effective sound-pressure value and the square of the reference sound pressure $p_0 = 20$ µPa:

$$L_p = 10 \log\left(\frac{{p_{eff}}^2}{{p_0}^2}\right) \text{ or}$$

$$L_p = 20 \log\left(\frac{p_{eff}}{p_0}\right).$$

The sound pressure level is given in decibels (dB). The frequency-dependent, A-weighted sound pressure level L_{pA} as measured at a distance of d = 1 m is frequently used to characterize sound sources.

Sound intensity level
The sound intensity level L_I is ten time the logarithm to the base 10 of the ratio of sound intensity and reference sound intensity $I_0 = 10^{-12}$ W/m^2:

$$L_I = 10 \log\left(\frac{I}{I_0}\right).$$

The sound intensity can be measured directly with a probe.

Interaction of two or more sound sources
If two independent sound fields are superimposed, their sound intensities or the squares of their sound pressures must be added. The overall sound level is then determined from the individual sound levels in accordance with Table 3.

Table 3: Overall sound level with superimposition of independent sound fields

Difference between 2 individual sound levels	Overall sound level = higher individual sound level + allowance of:
0 dB	3.0 dB
1 dB	2.5 dB
2 dB	2.1 dB
3 dB	1.8 dB
4 dB	1.5 dB
6 dB	1.0 dB
8 dB	0.6 dB
10 dB	0.4 dB

Measured quantities for noise immissions

Excessive noise
Excessive noise is the classification of a noise as an undesirable sound event and is dependent on the following factors:
– on the noise itself, measurable by physical variables (e.g. frequency, sound pressure level, or sound power),
– on the subjective attitude to the noise of the person affected,
– on the personal state of the person affected, and
– on the concrete situation in which the noise occurs.

Protection against excessive noise and in particular environmental noise is gaining increasing importance as an environmental theme. The recording and evaluation of noise immissions forms the basis for efficient noise reduction.

Rating sound level
The effect of noise on a human being is evaluated using the rating sound level L (see also DIN 45645-1 [5]; an ISO standard is currently being worked on). This is a measure of the mean noise immission over a period of time (e.g. eight working hours). With fluctuating noises it is either measured directly with integrated measuring instruments or calculated from individual sound-pressure-level measurements and the associated periods of time of the individual sound effects (see also DIN 45641, [6]). Noise-immission parameters such as pulsation (short strong deviations of the sound level from the mean level) and tonal quality (heavy dominance of one or more discrete frequencies) can be taken into account through level allowances.

Energy-equivalent continuous sound level
In the case of noises which fluctuate over time, the mean A-weighted sound pressure level resulting from the individual sound pressure levels and the individual exposure times equals the energy-equivalent continuous sound level L_{Aeq} (see also DIN 45641).

Table 4 provides guide values for the rating sound level (Germany; TI Noise [7]), measured outside the nearest residential building (0.5 m in front of an open window).

Noise indices
The EU Directive 2002/49/EC [8] defines the noise indices binding on the EU L_{den} and L_{night} as standard descriptors for offending noise over a full day (L_{den} for day, evening, night) and a night (L_{night}) along similar lines to the above-described L_{Aeq}. The assessment period here is one calendar year. Crashes for evening (5 dB) and night (10 dB) are taken into consideration for L_{den} in contrast to L_{Aeq}.

$$L_{den} = 10 \log \frac{1}{24} \cdot \left(12 \cdot 10^{\frac{L_d}{10}} + 4 \cdot 10^{\frac{L_e+5}{10}} + 8 \cdot 10^{\frac{L_n+10}{10}}\right)$$

where L_d, L_e and L_n are the values for day, for evening, and for night.

Sound exposure level
The sound exposure level SEL is used to evaluate individual excessive-noise events (e.g. an airplane starting) in sensitive conservation areas. For the purpose of calculation the sound energy of the complete event is recorded and its energy is then distributed to 1 second. It is therefore done as though the event always happens within a period of one second. The sound level is then calculated on this basis.

Table 4: Guide values for permissible noise pollution as per TI Noise [7]

	Day	Night
Purely industrial areas	70 dB (A)	70 dB (A)
Areas with predominantly commercial/industrial premises	65 dB (A)	50 dB (A)
Mixed areas	60 dB (A)	45 dB (A)
Areas with predominantly residential premises	55 dB (A)	40 dB (A)
Purely residential areas	50 dB (A)	35 dB (A)
Health resorts, hospitals, etc.	45 dB (A)	35 dB (A)

Perceived sound levels

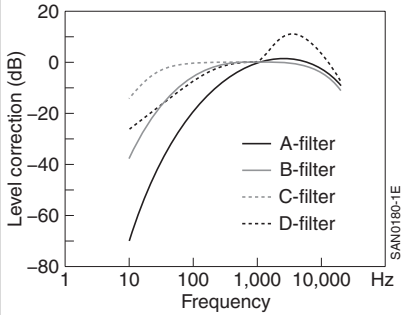
The human ear can distinguish approximately 300 levels of acoustic intensity and 3,000 to 4,000 different frequencies (pitch levels) in rapid temporal succession and evaluate them according to complex patterns. However, there is not necessarily any direct correlation between perceived loudness levels and (energy-oriented) technically defined sound levels. Corrections (A-, B- and C-weighting) take into account the dependence on frequency of human hearing (Figure 1).

The phon unit and the definition of loudness in "sone" provide a rough approximation of subjective sound-level perception. Sound-level measurements alone do not suffice to define the nuisance and disturbance potential of noise emanating from machinery and equipment. A hardly perceptible ticking noise can thus be perceived as extremely disturbing, even in an otherwise loud environment.

Loudness level

The loudness level L_S is a comparative measure of the intensity of the subjective perception of a sound, measured in "phon". The loudness level of a sound (pure tone or noise) is the sound pressure level of a standard pure tone which, under standard listening conditions, is judged by a normal observer to be equally loud. The standard noise level is a plane sound wave at a frequency of 1,000 Hz impinging on the observer's head from the front.

Figure 1: Frequency-dependent correction values for the sound level to take into account human hearing



This is known internationally as the "loudness level". A difference of 8 to 10 phon is perceived as twice or half as loud.

Phon

The standard pure tone judged as being equally loud has a specific value in dB. This value is given as the loudness level of the tested sound, and has the designation "phon". As human perception of sound is frequency-dependent, the dB values of the tested sound for notes, for example, do not agree with the dB values of the standard pure tone. The connection between curves of equal sound-level perception and sound pressure in decibels is always empirically determined. These curves were published for the first time in 1933 by Fletcher and Munson. They provided the basis for the isophone curves in accordance with DIN ISO 226 [9] used today. Figure 2 shows curves based on this standard.

Loudness in "sone"

The loudness S is the measure used to define subjective noise levels. The starting point for defining the sone is: How much higher or lower is the perceived level of a particular sound relative to a specific standard. Definition: Loudness level of L_S = 40 phon corresponds to a loudness S = 1 sone. Doubling or halving the loudness is equivalent to a variation in the loudness level of approx. 10 phon.

There is a DIN loudness standard for calculating stationary sound using tertiary levels (Zwicker method [10]). This procedure takes into account both frequency weighting and the screening effects of hearing.

Pitch, sharpness

The spectrum of perceptible sound can be divided into 24 hearing-oriented frequency groups (bark). The Bark scale is a psychoacoustic scale for perceived pitch. The loudness/pitch distribution can be used to quantify other subjective aural impressions – such as the sharpness (e.g. metallic harmonic distortion) of a noise.

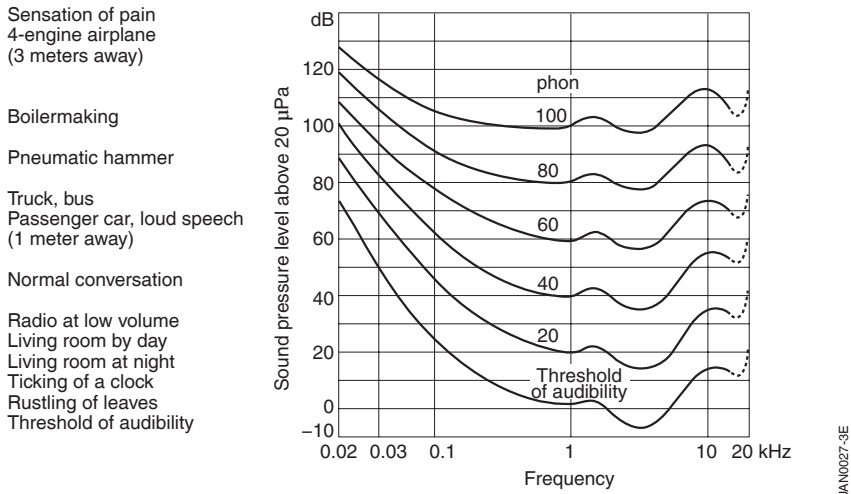
Articulation index

The basic prerequisite for recognizing speech is the correct transmission wherever possible of the sound waves of what is spoken from the sender (mouth) to the receiver (ear). The articulation index (AI) offers a computing model for predicting speech intelligibility in the event of interference noise. It is assumed here that the speech information is distributed to the different frequency bands of the acoustic signal. Each band then makes a contribution to speech intelligibility. If a band has a signal-to-noise ratio of more than 15 dB, the speech component in this band is evaluated as being intelligible. The sum total of all the bands with this noise ratio is multiplied by their specific weighting factor to give the articulation index AI in %.

References

[1] DIN 1320: Acoustics – Terminology.
[2] DIN EN ISO 80000-8: Quantities and units – Part 8: Acoustics.
[3] DIN EN 61260: Electroacoustics – Octave-band and fractional-octave-band filters.
[4] B. Lenze: Einführung in die Fourier-Analysis. 3rd Ed., Logos Verlag, Berlin, 2000.
[5] DIN 45645-1: Determination of rating levels from measurement data – Part 1: Noise immission in the neighbourhood.
[6] DIN 45641: Averaging of sound levels.
[7] Sixth General Administrative Regulation on the Federal Immission Protection Law (Technical Instructions on Noise Abatement – TI Noise) of 26 August 1998 (GMBl No. 26/1998 P. 503).
[8] Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise.
[9] DIN ISO 226: Acoustics – Normal equal-loudness-level contours.
[10] DIN 45631: Calculation of loudness level and loudness from the sound spectrum – Zwicker method.

Figure 2: Isophone curves in acc. with DIN ISO 226 with an allocation of noises (examples)



Optical technology

Electromagnetic radiation

Electromagnetic radiation denotes the wave nature of visible light, but also of radiation that is not visible to the human eye (e.g. ultraviolet light, X-radiation, radiation in the infrared range). Thus, electromagnetic radiation is subdivided by reference to its wavelength λ (Table 2).

Table 1: Quantities and units

Quantity		SI unit
A	Area A_1 Radiating area A_2 Illuminated area	m^2
E_e	Irradiance	W/m^2
I_e	Radiant intensity	W/sr
L_e	Radiance	$W/(m^2 sr)$
M_e	Radiant excitance	W/m^2
P_e	Power	W
Q_e	Radiant energy	Ws
R	Reflection coefficient	%
H_e	Irradiation	Ws/m^2
$K(\lambda)$	Absolute spectral sensitivity	lm/W
$V(\lambda)$	Spectral luminous efficiency	–
Φ_e	Radiant flux	W
r	Distance	m
t	Time	s
ε_1	Angle, incident ray (to normal of a surface)	°
ε_2	Angle, refracted ray	°
ε_3	Angle, reflected ray	°
$\varepsilon_{1,max}$	Angle of total reflection	°
η	Luminous efficiency	lm/W
Ω	Solid angle	sr
λ	Wavelength	nm
E_{phot}	Photon energy	eV
h	Planck's quantum of action	Js
c	Speed of light in a vacuum	m/s
n	Refractive index	–

Geometric optics

The description of optical systems can be limited to two models – geometric optics and wave optics. Geometric optics (also called ray model) describes the propagation of light and its deflection by imaging elements with simple geometric considerations. But this is only permitted provided the imaging elements (e.g. mirrors, lenses) are much larger than the wavelength of light.

Effects such as interference, diffraction and polarization are described in physics by wave optics. This is concerned with the propagation of light in the form of a wave.

Basic principles of geometric optics
A large part of the imaging systems used in optics can be described by geometric optics. Radiation propagation is explained by "light beams" and can be described by means of simple geometric laws.

Table 2: Ranges of electromagnetic radiation

Designation	Wavelength range
Gamma radiation	0.1…10 pm
X-radiation	10 pm to 10 nm
Ultraviolet radiation	10…380 nm
Visible radiation	380…780 nm
Infrared radiation	780 nm to 1 mm
Millimeter waves (EHF)	1…10 mm
Centimeter waves (SHF)	10…100 mm
Decimeter waves (UHF)	100 mm to 1 m
Ultrashort waves (VHF)	1…10 m
High-frequency waves (HF)	10…100 m
Medium waves (MF)	100 m to 1 km
Long waves (LF)	1…10 km
Myriameter waves (VLF)	10…100 km

Basic axioms of geometric optics
Fermat's principle is the basic principle for geometric optics. The basic axioms of geometric optics can be derived from the basic statement of this principle to the effect that light always chooses the shortest path between two points [1]:
– Light beams are straight in homogenous media.
– At the boundary between two homogenous materials the beams are deflected according to the law of reflection and the law of refraction.
– Each beam path is reversible.
– Light beams can cross without influencing each other.

Law of refraction
Snell's law of refraction [2] can also be derived from Fermat's principle. This law describes the transition of a light beam from one medium (e.g. air) to another medium (e.g. glass). At the interface of the two media the light is deflected due to the change of refractive index. An incident ray is split into a refracted ray and a reflecting ray (Figure 1). The refracted ray is governed by the law of refraction:

$$n_1 \sin \varepsilon_1 = n_2 \sin \varepsilon_2 .$$

For vacuum and dielectric media (e.g. air, glass, plastics) the refractive indices n are real numbers (Table 3). Materials with a large refractive index n are referred to as optically thick, materials with a small refractive index as optically thin.

The characteristic of refractive indices of being dependent on the wavelength is called dispersion. In most cases they decreases as the wavelength increases.

Law of reflection
The following formula applies to the direction of the reflecting ray (law of reflection):

$$\varepsilon_3 = \varepsilon_1 .$$

Figure 1: Refraction of a light beam at the interface between two media with different refractive indices
a Medium 1 with refractive index n_1,
b Medium 2 with refractive index n_2.
1 Incident ray,
2 Refracted ray,
3 Reflected ray.
ε_1 Angle, incident ray,
ε_2 Angle, refracted ray,
ε_3 Angle, reflected ray.

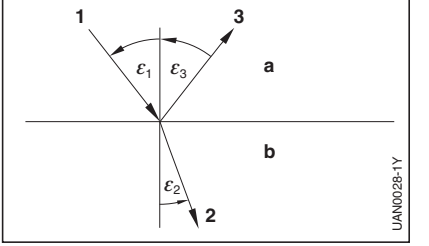
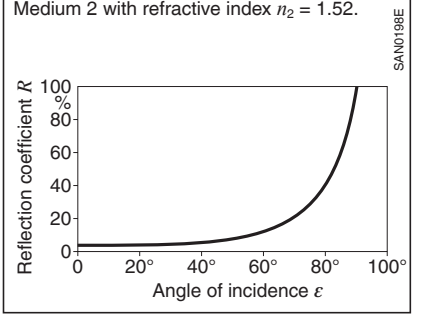


Table 3: Refractive index n of some media
(For yellow light of wavelength λ = 589.3 nm)

Vacuum, air	1.00
Ice (0 °C)	1.31
Water (20 °C)	1.33
Quartz glass	1.46
Polymethylmethacrylate	1.49
Standard glass for optics (BK 7)	1.51673
Window glass	1.52
Glass for headlamp lenses	1.52
Polyvinylchloride	1.54
Polycarbonate	1.58
Polystyrene	1.59
Epoxy resin	1.60
Gallium arsenide (depending on doping)	approx. 3.5

Figure 2: Dependence of the reflection coefficient on the angle of incidence
Medium 1 with refractive index n_1 = 1.00,
Medium 2 with refractive index n_2 = 1.52.



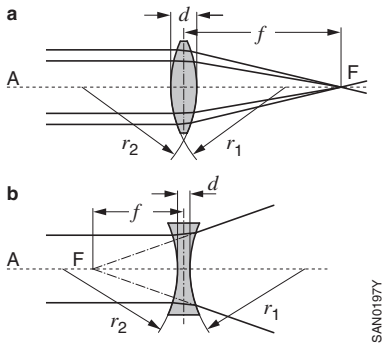
The angle of reflection is therefore equal to the angle of incidence. The intensity ratio of reflected to incident radiation (degree of reflection, reflection coefficient) is described by the Fresnel formula [3] and is dependent on the angle of incidence ε_1 and on the refractive indices of the adjoining media (Figure 2). In the transition from air (n_1 = 1.00) to glass (n_2 = 1.52) the upshot is that, in the case of vertical ray incidence (ε_1 = 0), 4.3 % of the ray energy is reflected.

Total reflection

In the transition from an optically thicker to an optically thinner medium ($n_1 > n_2$) the angle of refraction ε_2 attains for a specific angle of incidence ε_1 its highest possible value (90°). This angle of incidence $\varepsilon_{1,max}$ is called the critical angle for total reflection. According to the law of refraction, the following is obtained for total reflection:

$$\sin \varepsilon_{1,max} = \sin 90° \frac{n_2}{n_1} = \frac{n_2}{n_1} .$$

Figure 3: Beam path for spherical lenses
a) Convex lens,
b) Concave lens.
A Optical axis,
F Focal point.
f Focal length,
d Thickness of lens,
r_1, r_2 Bending radii of lens.



Optical components

Spherical lenses

The optical effect and the imaging property of a lens are determined by the shape of the interfaces and by the refractive index of the lens material [4]. The imaging properties of light beams can be determined using Snell's law of refraction.

The functioning principle of a lens is divided into two subgroups based on its shape (Figure 3). Lenses that bring light to a focal point are called convergent lenses (convex lenses). In contrast, parallel beams of rays are spread out by divergent lenses (concave lenses).

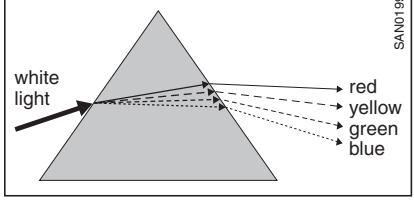
Prisms

In the case of a prism, the dependence of the refractive index on the wavelength of the incident light is utilized (dispersion). Thus, blue light (λ = 420 to 490 nm) is refracted at a different angle ε_2 from red light (λ = 650 to 750 nm). When white light shines onto a prism, it is broken down into different component parts (color components) (Figure 4).

This effect also appears in rainbows. The different color components of sunlight are refracted differently on water droplets in the air.

Vehicle headlights use lenses (elements comprising cylindrical lenses and prisms) to favorably influence the radiation coming from the reflector.

Figure 4: Beam path for a prism



Reflectors
Reflectors (mirrors) are used in optics to divert light in a particular direction (Figure 5). Reflectors are used in among other things headlights, where they specifically direct and concentrate light.

The function of reflector is to capture light from the headlight bulb, to achieve as great a range as possible, and to influence the distribution of the light on the road in such a way as to satisfy the legal requirements. Design (e.g. when installed in the bumper) places additional demands on the headlights.

Whereas previously almost exclusively paraboloids were used for reflectors, the above-mentioned, partly contradictory requirements can today sometimes only be satisfied by stepped reflectors, free-form surfaces or new headlight concepts (see Lighting equipment).

Basically speaking, the larger the lens aperture area, the greater the headlight range that can be achieved. On the other hand, the luminous efficiency increases with the size of the solid angle captured by the reflector.

Color filters
For special applications, e.g. lights and lamps on vehicles, there are precise regulations regarding the wavelength of the light used to suit the specific purpose (turn-signal lamp, stop lamp). Color filters are used here to attenuate or suppress unwanted spectral ranges.

Wave optics

The previous considerations only took into account macroscopic systems, i.e. the imaging elements considered were much larger than the wavelength of the observed light beam. Wave optics is concerned with systems in which light is described as an electromagnetic wave. The color of the radiation is defined by the wavelength. Thus, monochromatic light is made up only of light waves of one wavelength, whereas white lights contains light waves of different wavelengths.

Polarization
Polarization is also taken into account when a system is considered from the wave-optics standpoint. This describes the orientation of the waves or the oscillations in relation to the plane of incidence. There are three different types of polarization: linear, circular and elliptical. In the case of linear polarization, the amplitude of the electric field varies with a constant direction of propagation; in the case of circular polarization, the direction varies while the amplitude of the electric field remains constant. Elliptical polarization describes a mixed form of linear and circular polarization.

Linear polarization is further subdivided into s-polarization (perpendicular to the plane of incidence) and p-polarization (in the plane of incidence).

As shown in Figure 6, the characteristic of the Fresnel equations calculated in Figure 2, taking into account s- and

Figure 5: Beam path for a parabolic reflector
A Optical axis,
F Focal point (focus),
P Vertex of parabola.

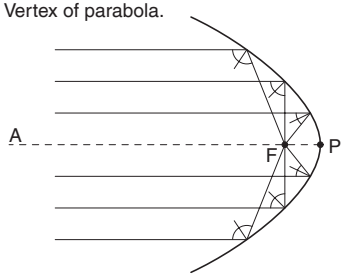
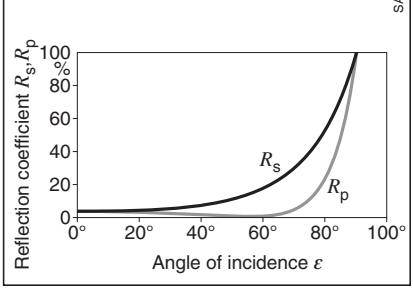


Figure 6: Reflection coefficient
Medium 1 with refractive index n_1 = 1.00,
Medium 2 with refractive index n_2 = 1.50.



p-polarization, is divided into two different curves. The plane perpendicular to the plane of propagation is called the plane of oscillation. It follows that only waves whose electric field components oscillate in the plane of oscillation are polarizable. These are also called transversal waves (light). In contrast, longitudinal waves (e.g. sound) oscillate in the direction of propagation and are therefore not polarizable.

Interference

The superposition of two waves of equal wavelength can lead to interference. According to the superposition principle, places of constructive and places of destructive interference are formed here. An interference pattern with interference maxima and interference minima is produced. An established example of this is Young's doubles-slit experiment [4], which proved for the first time the wave nature of light.

Figure 7: Intensity curve $I(x)$ for diffraction at the single slit

L Lens, B Aperture,
D Opening diameter,
x Distance,
$I(x)$ Intensity profile.

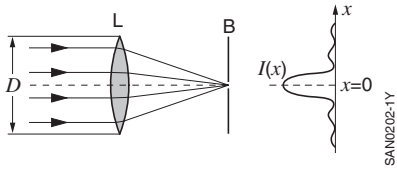
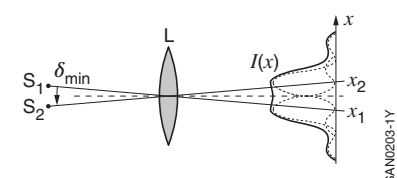


Figure 8: Resolving capacity limited by diffraction of an optical system

L Lens,
S_1, S_2 Point sources,
x_1, x_2 Imaging of point sources,
δ_{min} Minimum resolving angle,
$I(x)$ Intensity profile.



Diffraction

Diffraction phenomena of light waves can also be described using wave optics. The term diffraction is used to refer to the situation where an optical wave encounters an obstacle (e.g. aperture, single slit). The light wave is deflected at the obstacle and new elementary waves are created in accordance with Huygens'principle [4]. These waves interfere with each other and a diffraction pattern is produced in response to interference effects. Thus light penetrates into areas which should be dark according to geometric optics.

Diffraction at a single slit produces in response to interference effects the intensity curve $I(x)$ shown in Figure 7.

When diffraction is considered at an aperture plate or at a circular aperture, the resulting electric field after the aperture can be described by the Bessel functions of the first kind $J_1(2\pi\, r)$ [5]:

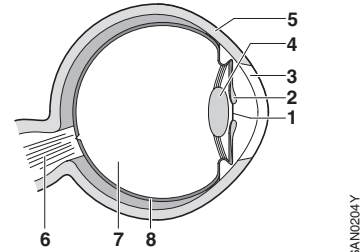
$$E(r) = \frac{E_0}{\pi r} J_1(2\pi\, r)\,.$$

The resulting intensity distribution after diffraction at the aperture is obtained from

$$I(r) \sim E^2(r).$$

Figure 9: Structure of the eye

1 Pupil, 2 Iris, 3 Cornea, 4 Lens,
5 Sclera, 6 Optic nerve,
7 Vitreous humor, 8 Retina.



Assuming that only very weak (especially the radius) secondary maxima develop as a result of diffraction at a circular aperture, the magnitude of the principal maxima can be determined from the zero point of the Bessel function. This is obtained at $r = 0.6098$.

If diffraction is disregarded, optical systems can achieve an infinite resolution. Diffraction significantly reduces the resolving capacity of any optical system (e.g. lenses, objectives). Resolving capacity U denotes the smallest still perceptible distance between two point sources (S1 and S2).

Basically, optical elements are limited by a circular aperture. The resolving capacity of any optical system can thus be calculated from the size of the diffraction pattern of a circular aperture. The intensity curve $I(x)$ is obtained for each point source. The minimum angle δ_{min} up to which the points on the screen can still be perceived separately (Figure 8) is calculated as described in the following equation from the wavelength λ of the observed radiation and the opening diameter D:

$$\sin \delta_{min} = 1.22 \frac{\lambda}{D}.$$

Since the diameter of the aperture opening is calculated from the double radius, the factor 1.22 was determined from the zero point of the Bessel function. The resolving capacity of the system can be determined from δ_{min} at $U = 1/\delta_{min}$.

The human eye can also be considered to be an optical system. In this case the focal plane is the pupil and the shown intensity profile $I(x)$ of a point light source is imaged on the retina (Figure 9). The opening diameter D is determined by the size of the pupil. For a wavelength of $\lambda = 550$ nm and a pupil size of 1.5 mm (daylight vision) $\delta_{min} = 0.02°$ is obtained for the minimum angle. From this, at a distance of 25 cm to the eye, the minimum resolvable distance between S_1 and S_2 is calculated at 0.11 mm.

Lighting quantities

Light sources have – on the basis of their nature – radiation of different intensities and different wavelengths. The light source is classified by reference to the characteristic wavelength into the wavelength spectrum.

Physical characteristic quantities have become established for the change in intensity of the light source and to describe the effect of light on human beings. A distinction is made here between radiometric and photometric characteristic quantities. In photometry the human eye is evaluated as a detector. This is confined to the ultraviolet (UV) and the visible spectral ranges. In radiometry the power of the light source measured by a detector is determined, thereby also providing measurements in the infrared and ultraviolet ranges and with gamma rays. The lighting quantities in radiometry (radiation quantities) are identified for improved intelligibility by the indices “e” (“e” for energy).

Radiation quantities

Radiant energy Q_e

The radiant energy Q_e comprises the total energy of the light wave.

Radiant flux Φ_e

Radiant flux Φ_e denotes the amount of energy dQ_e that is transported per time unit dt by the light wave:

$$\Phi_e = \frac{dQ_e}{dt}.$$

Irradiance E_e

The proportion of the radiant flux $d\Phi_e$ which strikes a defined surface area dA is called irradiance E_e:

$$E_e = \frac{d\Phi_e}{dA}.$$

The totality of the radiated energy that strikes a bright surface element dA is determined.

Radiant intensity I_e
Radiant intensity I_e denotes the radiant flux $d\Phi_e$ emitted by a point wave in a particular direction (solid angle $d\Omega$).

$$I_e = \frac{d\Phi_e}{d\Omega}.$$

Radiance L_e
Radiance L_e denotes the radiant intensity on a vertical surface.

$$L_e = \frac{dI_e}{dA\cos\varepsilon} = \frac{d^2\Phi_e}{d\Omega\, dA\cos\varepsilon}.$$

Radiant excitance M_e
Radiant excitance M_e denotes the radiant flux $d\Phi_{e,H}$ of a light source radiated into the half-space:

$$M_e = \frac{d\Phi_{e,H}}{dA}.$$

Irradiation H_e
Irradiation H_e denotes the proportion of the radiant energy dQ that strikes per time dt a surface element dA.

$$H_e = \int E_e\, dt.$$

Lighting quantities in photometry
Spectral luminous efficiency $V(\lambda)$
Defining the lighting quantities in photometry does not take into account the fact that the eye does not have constant spectral luminous efficiency. The radiation visible to the human eye is in the wavelength range of 380 nm (blue) through 780 nm (red). The eye is at its most sensitive in daylight in the yellow-green range around 555 nm; in weaker light conditions this figure shifts to lower wavelengths. A greater radiant energy is required to achieve the impression of equal brightness for other wavelengths. Spectral luminous efficiency $V(\lambda)$ is defined as the ratio of radiant energy at 555 nm to the radiant energy for the different wavelengths.

Because not all human eyes are identical, a non-dimensional, standardized luminous-efficiency function of the human eye $V(\lambda)$ was defined in DIN 5031-3 [9] for lighting measurements and calculations. This was determined for different wavelength ranges and for daylight and nightlight conditions (Figure 10).

Luminous efficacy of radiation
The absolute spectral efficiency $K(\lambda)$, also called the luminous efficacy of radiation, is the quotient from the physiological quantity luminous flux Φ and the physical radiant flux Φ_e. For daytime vision (eye adapted to the light) the maximum value $K_{m,T}$ of $K(\lambda)$ is obtained at a wavelength of λ = 555 nm. For nighttime vision (eye adapted to the dark) the maximum $K_{m,N}$ is obtained at λ = 505 nm. The following relations are obtained for the absolute spectral efficiency $K(\lambda)$:

Figure 10: Luminous-efficiency function of the human eye
$V(\lambda)$ Day-value curve,
$V'(\lambda)$ Night-value curve.

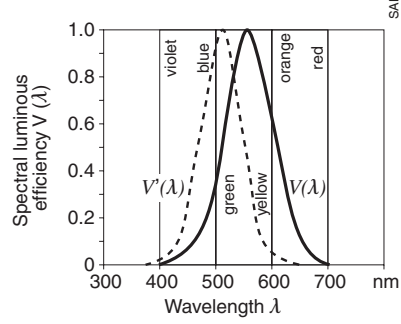


Table 4: Comparison of radiation quantities and photometric quantities

Radiation quantities		Photometric quantities	
Designation	Unit	Designation	Unit
Radiant energy Q_e	Ws	Quantity of light Q	lm s
Radiant flux Φ_e	W	Luminous flux Φ	lm
Radiant intensity I_e	W/sr	Luminous intensity I	cd = lm/sr
Radiance L_e	W/(m^{2}sr)	Luminance L	cd/m^2
Irradiance E_e	W/m^2	Illuminance E	lx = lm/m^2
Radiant excitance M_e	W/m^2	Luminous excitance M	lx
Irradiation H_e	Ws/m^2	Lumination H	lx s

Daytime vision:
$K(\lambda) = K_m\, V(\lambda)$,
$K_{m,T}$ = 683 lm/W.

Nighttime vision
$K'(\lambda) = K_m\, V'(\lambda)$,
$K_{m,N}$ = 1,699 lm/W.

$K(\lambda)$ enables the radiation quantities X_e and the photometric quantities X to be linked (Table 4). The following applies:

$$X = K_m \int X_{e\lambda}\, V(\lambda)\, d\lambda,$$

$$= \int X_{e\lambda}\, K(\lambda)\, d\lambda,$$

where $X_{e\lambda} = \frac{dX_e}{d\lambda}$.

Luminous flux Φ
The luminous flux Φ denotes the weighting of the radiometric radiant flux Φ_e with the wavelength-dependent efficiency curve $K(\lambda)$ of the human eye. The luminous flux Φ determines the light output radiated into the totality of solid angles. All further photometric quantities are linked with the luminous flux.

Quantity of light Q
The quantity of light Q is the photometric equivalent to the radiometric radiant energy Q_e. It is calculated from the integral of the luminous flux Φ over a certain time dt:

$$Q = \int \Phi(t)\, dt.$$

Luminous intensity I
The luminous flux Φ that is radiated into a certain solid angle Ω is defined as the luminous intensity I. It offers a measure of the luminous emittance of a source radiated in a certain direction and is calculated from the radiometric radiation intensity:

$$I = \frac{\Phi}{\Omega}.$$

Illuminance E
In contrast to luminous intensity I, illuminance E denotes the luminous flux of a light source radiated to a certain surface A:

$$E = \frac{\Phi}{A}.$$

Luminance L
Luminance L defines the brightness sensation in the human eye caused by an illuminated (or also self-luminous) surface. It is obtained from the radiometric quantity for the radiant intensity:

$$L = \frac{I}{\Omega}.$$

Luminous excitance M
The proportion of the luminous flux $d\Phi$ that is generated by a defined surface element dA is known as luminous excitance. It is the photometric counterpart to radiometric radiant excitance:

$$M = \frac{d\Phi}{dA}.$$

Lumination H
The lumination H is calculated from the illuminance E_V that strikes the surface dA over a period of time dt.

$$H = \int E_V(t)\, dt.$$

Further terms
The lighting quantities are supplemented by the general terms explained in the following.

Luminous efficiency η
Luminous efficiency η is calculated from the ratio of emitted luminous flux Φ to power input P:

$$\eta = \frac{\Phi}{P}.$$

Luminous efficiency is a measure of the efficacy of converting electric power P into luminous flux Φ. It cannot exceed the maximum value of the luminous efficacy of radiation K_m = 683 lm/W for the wavelength λ = 555 nm.

Solid angle Ω
The solid angle Ω describes the ratio of the penetrated section of the spherical surface (of a sphere concentric to the radiation source) to the square of the sphere radius. The total surface of the sphere is $4\pi r^2$, thus the full solid angle is

$\Omega = 4\pi$ sr (sr = steradian).

Steradian is a non-dimensional quantity for the solid angle.

Contrast
The contrast determines the luminance ratio between two neighboring surfaces or the maximum difference in intensity within an illuminated (or self-luminous) surface. The contrast describes the ratio of luminance or the illuminance of a bright area of an image to a dark area of the same image. According to the standard ISO IEC 21118 [10] the contrast must be determined from the values of a black-and-white checkerboard pattern (consisting of sixteen equal squares).

The small-area contrast for alternating black and white lines can also be determined. This is a measure for the specification of an optical lighting system to shown fine details on a screen. Both vertical and horizontal lines are measured for this purpose.

Table 5: Examples of some laser types

Laser type	Wavelength	Applications
He-Ne laser	632 nm	Metrology, holography
CO_2 laser	10.6 µm	Material processing
ND: YAG laser	1,064 nm	Material processing
Semiconductor laser	e.g. 670 nm e.g. 1,300 nm	Measuring technology Telecommunications
Ytterbium fiber laser	e.g. 1,070 nm	Material processing

Laser technology

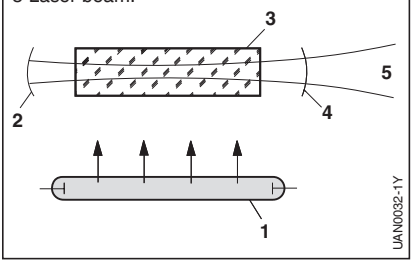
Compared to other light sources, the laser (Light Amplification by Stimulated Emission of Radiation) has the following characteristic properties:
– Monochromatic radiation, i.e. a limited wavelength range,
– Very high radiation density,
– Low beam expansion,
– Good focusability,
– High time and spatial coherence of radiation.

Functioning principle
A laser contains a laser-active medium that can be gaseous, solid or liquid (Table 5). The supply of energy can place the atoms or the molecules of the active medium in an excited state (Figure 11). This process is called pumping and can take place electrically (by the application of a voltage) or optically (with another light source).

After a certain period of time the excited particles of the active medium relax back into their initial state and in so doing dissipate their energy through the spontaneous emission of a photon (light particle). The energy of the photon is determined by the quantized energy states of the active medium and stipulates the wavelength λ of the laser light:

$$E_P = \frac{hc}{\lambda}.$$

Figure 11: Laser principle
1 Pumping light source,
2 Resonator mirror,
3 Laser-active material,
4 Partly transparent mirror,
5 Laser beam.



h is Planck's quantum of action and c the speed of light in a vacuum. For h and c the following values apply:
$h \approx 6.62606957 \cdot 10^{-34}$ Js
$c \approx 300{,}000{,}000$ m/s

The resonator consists of mirror surfaces at the ends of the active medium and causes the spontaneously emitted photons to be reflected back into the active medium. As they pass through the active medium again they cause new photons of equal wavelength and identical phase position to be emitted. This is called stimulated emission.

The resonator is responsible for radiation amplification and for the desired beam characteristic. The laser beam emerges at the end of the resonator via a semi-transparent mirror.

Depending on the active medium and the laser type, lasers can be operated with continuous-wave radiation or in pulsed mode with pulse lengths down to below 1 fs (10^{-15} s) [12, 13].

Areas of application
Laser measuring technology permits the noncontact, non-interacting testing of production tolerances of superfinished surfaces (e.g. fuel injectors). Resolutions in the nm range are achieved using interferometric methods.

Lasers facilitate high-precision, flexible and high-speed material processing/machining in production engineering. For example, hole diameters of 30 µm can be achieved with laser drilling.

Further laser applications in technology are holography (spatial image information), automatic character recognition (bar-code scanners), information recording (CD scanning, 3D space surveying), material processing/machining, microsurgery, and transmitters for data transmission in optical waveguides.

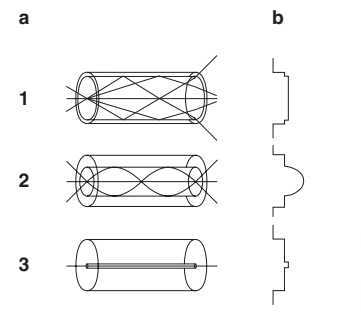
Specific regulations are to be observed when handling laser products. Laser products are classified according to potential hazards. For details, refer to DIN EN ISO 60825 [11].

Optical fibers/waveguides

Design
Optical fibers/waveguides transmit under controlled conditions electromagnetic waves in the ultraviolet (UV), visible, and infrared (IR) ranges of the spectrum. They are made of quartz, glass, or polymers, usually in the form of fibers or channels created in transparent materials with a core whose refractive index is usually higher than that of the cladding. Thus, light launched into the core is retained in that area by means of refraction or total reflection and routed.

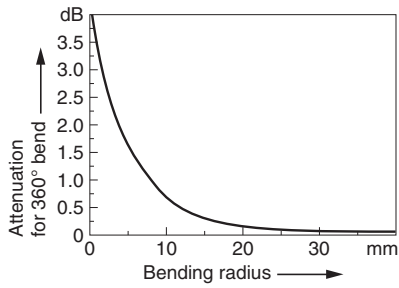
Depending on the refractive index profile, a distinction is made between four types of fiber (Figure 12):
– The step-index fiber, with a sharply-defined boundary between the core and the cladding
– The graded-index fiber, with parabolic refractive index profile in the core
– The monomode fiber with a very small core diameter
– The photonic-crystal fiber with air-filled capillaries arranged periodically around the core. The arrangement corresponds to the step-index or monomode fiber.

Figure 12: Light propagation in optical fibers
a) Fibers schematically represented,
b) Refractive-index profile.
1 Step-index fiber,
2 Graded-index fiber,
3 Monomode fiber.



Step-index and graded-index fibers are multimode fibers, i.e. various oscillation modes of the light waves can be propagated along then at different speeds. The different oscillation modes of the light wave are dependent on the geometry and the material properties of the substrate material. Polymer fibers are always step-index fibers. Monomode fibers are, thanks to their geometry and the material used, designed in such a way that only the fundamental mode of the light wave can be propagated. Depending on the structure configuration, photonic-crystal fibers guide one or more modes.

Figure 13: Attenuation of a 360° bend depending on the bending radius
Source: [14].



Properties
Glass optical fibers have a high degree of transparency in the range from ultraviolet to infrared. Losses occur primarily due to the contamination of the fiber material used in the manufacturing process. Thus, H_2O molecules absorbed from the ambient air generate absorption band in the spectrum of the optical fibers. At 950 nm, 1,240 nm and 1,380 nm the transparency of the optical fibers is reduced by the absorption of the H_2O molecules in the glass fiber. When the spectrally resolved absorption of optical fibers is considered, minima are to be seen at 850 nm, 1,310 nm and 1,550 nm, i.e. attenuation is particularly low for the wavelengths 850 nm, 1,310 nm and 1,550 nm. Synthetic fibers absorb above 850 nm and below 450 nm.

Optical fibers can only absorb light from a restricted angular range Θ. The numerical aperture A_N = sin (Θ/2) (Table 6) serves as the measure for this.

The differences in dispersion and propagation time of the various modes cause an increasing broadening of the light pulses as the length of the fiber increases, and thus restrict the bandwidth. In photonic-crystal fibers, it is possible by means of appropriate microstructuring of the core the influence the dispersion and the efficiency of nonlinear effects to achieve desired results.

Optical fibers can be used within the temperature range of –40 to 135 °C; special versions can even be used up to 800 °C.

Table 6: Characteristic data of optical fibers/waveguides

Fiber type	Diameter		Wavelength	Numerical	Attenuation	Bandwidth
	Core [µm]	Cladding [µm]	[nm]	aperture (A_N)	[db/km]	[MHz · km]
Step-index fiber						
Quartz, glass	50…1,000	70…1,000	250…1,550	0.2…0.87	5…10	10
Polymer	200…>1,000	250…2,000	450…850	0.2…0.6	100…500	<100
Graded-index fiber	50…100	100…500	450…1,550	0.2…0.3	3…5	200…10,000
Monomode fiber	3…10	100…500	850…1,550	0.12…0.21	0.3…1	2,500…10,000
Photonic-crystal fiber	1…35	250…200	300…2,000	0.1…0.8	0.2…2	≤160,000

Areas of application

The main area of application for optical fibers/waveguides is in data transmission. Synthetic fibers are preferred for use in the LAN (Local Area Network) field. Graded-index fibers are the most suitable for medium ranges. Only monomode fibers are used for long-distance data transmission. In fiber-optic networks, erbium-doped glass fibers serve as optical amplifiers. Here, the transported radiation can be boosted by additional optical pumping with a semiconductor source.

Optical fibers/waveguides are used in vehicles in the MOST bus. The required compliance of bending radii means that installation in motor vehicles is critical. If the bending radii are too small, the attenuation is too great (Figure 13, [14]).

Optical fibers are being increasingly used in motor-vehicle lamps and sensors. Fiber-optic sensors generate neither scatter fields nor sparks, and are themselves insensitive to that kind of disturbance. They are currently employed in potentially explosive environments, in medicine, and in high-speed trains (ICE).

Energy transport is at the forefront in the area of material processing with laser beams, in microsurgery, and in lighting engineering.

References

[1] H. Haferkorn: Optik – Physikalisch-technische Grundlagen und Anwendungen. 4th Ed., Wiley-VCH, 2002.
[2] P.A. Tipler, G. Mosca: Physik. 6th Ed.; Springer, 2009.
[3] W. Demtröder: Experimentalphysik 2 – Elektrizität und Optik. 6th Ed.; Springer, 2013.
[4] G. Litfin: Technische Optik in der Praxis. 3rd Ed.; Springer, 2005.
[5] W. Nolting: Grundkurs Theoretische Physik 5/2 – Quantenmechanik – Methoden und Anwendungen (Springer-Lehrbuch), 7th Ed., Springer-Verlag, 2012.
[6] E. Hecht: Optik. 5th Ed., Oldenbourg Wissenschaftsverlag, 2009.
[7] F. Pedrotti, L. Petrotti, W. Bausch: Optik – Eine Einführung. Reihe Prentice Hall, Markt+Technik Verlag, 1996.
[8] F. Pedrotti, L. Petrotti: Introduction to Optics. 3rd Ed., Pearson Education Limited, 2013.
[9] DIN 5031-3: Optical radiation physics and illuminating engineering; quantities, symbols and units of illuminating engineering.
[10] ISO IEC 21118: Information technology – Office equipment – Information to be included in specification sheets – Data projectors.
[11] DIN EN 60825: Safety of laser products.
Part 1: Equipment classification and requirements (2008).
Part 2: Safety of optical fibre communication systems (2011).
Part 4: Laser guards (2011).
Part 12: Safety of free space optical communication systems used for transmission of information (2004).
[12] W. Radloff: Laser in Wissenschaft und Technik. Spektrum Akademischer Verlag, 2011.
[13] F. K. Kneubühl, M. W. Sigrist: Laser. 7th Ed., Vieweg+Teubner, 2008.
[14] A. Grzemba (Editor): MOST – Das Multimedia-Bussystem für den Einsatz im Automobil. 1st Ed., Franzis-Verlag, 2007.

Thermodynamics

Basic principles

System

Thermodynamic systems are typically divided into three types. What all the systems have in common is that they are delimited from the surroundings by the system boundary within which the system is located.

Isolated system
The system boundary of an isolated system is impermeable to all types of heat Q, work W and mass m (e.g. the medium inside a perfectly insulated Thermos flask).

Closed system
A closed system allows the exchange of heat and work with the surroundings across the system boundary, but not of mass (e.g. a gas in a cylinder sealed by a moving piston).

Open system
Finally, in an open system heat, work and mass can be exchanged across the system boundary with the surroundings (e.g. an open cooking pot).

State and process variables

State variables
In thermodynamics states of systems and (changes to) processes are described by way of state variables such as for example temperature T, pressure p, volume V or mass m. The changes to these variables are called changes of state. The state of a system is described uniquely by means of the state variables. To describe the state of a system which undergoes a change of state from state 1 to state 2, it is therefore not necessary to know the path taken between initial and final states, but solely to know the relevant state variables.

Process and process variables
Generally speaking, the stringing together of changes of state is called a process. In contrast to state variables, the process variables required to describe the process depend on the path between initial and final states. Work and heat are thus process variables, the type and with it also the magnitude of the work input differ e.g. depending on process control.

Table 1: Symbols and units

Quantity		SI unit
A	Cross-sectional area	m^2
a	Thermal diffusivity	m^2/s
c	Specific heat capacity	J/(kg·K)
	c_p Isobaric (constant pressure)	
	c_v Isochoric (constant volume)	
E	Energy	J
e	Specific energy	J/kg
H	Enthalpy	J
h	Specific enthalpy	J/kg
k	Heat-transmission coefficient	$W/(m^2 \cdot K)$
m	Mass	kg
n	Polytropic exponent	–
p	Pressure	$Pa = N/m^2$
Q	Heat	J
$\dot{Q}$	Heat flow dQ/dt	W
R_m	Universal gas constant	J/(mol·K)
	$R_m \approx 8.3145$ J/(mol·K)	
R_i	Specific gas constant	J/(kg·K)
	$R_i = R_m/M$ (M molar mass)	
R_λ	Thermal-conduction resistance	K/W
S	Entropy	J/K
s	Length	m
T	Thermodynamic temperature	K
ΔT	Temperature difference	K
	$\Delta T = T_1 - T_2$	
U	Internal energy	J
u	Specific internal energy	J/kg
V	Volume	m^3
v	Specific volume	m^3/kg
W	Work	J
W_t	Technical work	J
t	Time	s
α	Heat-transfer coefficient	$W/(m^2 \cdot K)$
ε	Emissivity	–
κ	Isentropic exponent	–
λ	Thermal conductivity	W/(m·K)
ρ	Density	kg/m^3
ν	Kinematic viscosity	m^2/s
σ	Stefan-Boltzmann constant	W/m^2K^4
	$\sigma \approx 5.6704 \cdot 10^{-8}$ W/m^2K^4	

When a system runs through a sequence of processes and ends again in the initial state, this is called a cycle. Cycles are important for describing many technical applications (e.g. engines, power plants, power stations, air conditioners).

Intensive and extensive state variables
State variables can be further subdivided into intensive and extensive variables. When an existing system is subdivided into two parts, all the variables whose values are maintained in the two new systems are called intensive variables (e.g. temperature, pressure). Variables whose values change are called extensive variables (e.g. volume).

Often of interest are system properties which do not depend on the absolute value of a system. It can therefore be practicable to divide the extensive variables by the system mass. This results in specific state variables which retain their value when the system is subdivided (e.g. specific volume $v = V/m$ and its reciprocal value, density $\rho = 1/v$).

Forms of energy
As in classic mechanics, the state variables "kinetic energy" and "potential energy" are defined in thermodynamics. When mass m moves at constant speed c, it thus has the kinetic energy

$$E_{kin} = \frac{1}{2} m c^2.$$

A mass m that is under gravity at the acceleration of free fall g and height z has the potential energy

$$E_{pot} = m g z.$$

In addition to these forms of energy, a thermodynamic system has a further inherent type of energy, internal energy U. As well as the movement of the system, this type of energy furthermore takes into account changes within the system due for example to a change in temperature. The total energy of a system is then:

$$E_{tot} = U + E_{kin} + E_{pot}\ .$$

Work and heat
Work W is generally defined as the integral of the force F acting at a point of application on the system via the distance covered ds (from s_1 to s_2):

$$W_{12} = \int_{s_1}^{s_2} F\,ds\ .$$

The special case of volume-change work is crucially important in thermodynamics. This is defined as

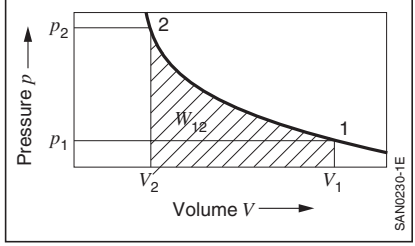
$$W_{12} = -\int_{V_1}^{V_2} p\,dV.$$

Volume-change work can be shown and read as the area (integral) under the curve in a pV diagram (Figure 1).

In contrast to work, heat Q is a form of energy that occurs in an unordered way. When energy is supplied to a system for example by an electric heater, the change in the system's state of equilibrium is caused by heat, but not by work, which in this example is equal to zero.

The two variables work and heat are counted positively when they are supplied to a system. When work or heat is rejected, they are negative. This equates to a system-egotistical standpoint.

Figure 1: Volume-change work as integral in the pV diagram



Laws of thermodynamics

Zeroth law
Two systems which are each in thermal equilibrium with a third system are also in thermal equilibrium with each other. All the systems thus have the same temperature.

First law
The first law of thermodynamics states that energy cannot be created or destroyed, but only converted from form to another, e.g. heat to work.

For an isolated system, in which there are no flows of mass or energy across the system boundaries, the sum of all the energy changes equals zero.

For a closed system, in which energy but not mass can flow across the system boundaries, the following applies:

$$Q_{12} + W_{12} = \Delta U + \Delta E_{kin} + \Delta E_{pot}.$$

The sum of the change in the internal energy ΔU and the kinetic energy ΔE_{kin}, and the potential energy ΔE_{pot} of the system is, therefore, the sum of the quantity of heat Q_{12} added and the work W_{12} done on the system (in the case of a change from state 1 to state 2).

In many technical applications a medium flows through an open system in a stationary flow process. It is therefore also necessary to consider that the intake and expulsion of the input or output mass is connected with work W_t and the input or output mass has an additional energy content. For a stationary flow process with an input and an output mass flow

$\dot{m}_1 = -\dot{m}_2 = \dot{m}$ the following applies:

$$\dot{Q}_{12} + \dot{W}_{t,12} = \dot{m}\,(\Delta h + \Delta e_{kin} + \Delta e_{pot}),$$

with the new state variable enthalpy:

$$h = u + pv.$$

e denotes the specific variable, i.e.:

$$e = \frac{E}{m}.$$

A perpetuum mobile of the first kind violates the first law of thermodynamics in that it creates “energy from nothing” or presupposes an (impossible) efficiency of more than 100 %.

Second law
Once the retention of the total energy has been described by the first law, the second law of thermodynamics specifies the direction of the energy exchange and the sequences of processes. For example, heat transfer can only take place from a hot body to a cold body, never the other way round. The decisive state variable is entropy. For reversible changes of state the change dS in the entropy S of a system is defined as:

$$dS = \frac{dQ}{T}.$$

In other words, the transferred quantity of heat dQ is divided by the temperature T at the point of heat exchange. If therefore a quantity of heat dQ is isothermally added for example to a system, this results in an increase in the statistical fluctuation of the kinetic energy of the atoms (effectively the addition of heat) and also an increase in its Entropy by the value dS.

For all irreversible changes of state and thus for all real technical processes (accompanied by loss) it follows as a result of energy dissipation that entropy production is positive. In the case of a reversible process, however, no entropy is created. To summarize, it can therefore be stated that the rate of entropy production is never negative and thus:

$$dS \geq 0.$$

A perpetuum mobile of the second type violates the second law of thermodynamics, while it can perfectly comply with the first law. The second main law states that it is not possible to convert heat into work when the ambient temperature stays the same. Nor is it possible to convert a quantity of heat which exists at a temperature level above the surroundings fully into work.

Third law
The third law assigns to the entropy at absolute zero, at which the temperature T = 0 K, a value that is not dependent on pressure, temperature, volume, etc. This is S_0 = 0 J/K.

It also states that absolute zero can be approached by a series of different processes only asymptotically, but never reached.

Table 2: Specific heat capacity c_p at constant pressure for some gases
Values apply to 1 bar and 293.15 K [1].

Gas	c_p [kJ/kg·K]
Nitrogen (N_2)	1.041
Oxygen (O_2)	0.9189
Helium	5.251
Air	1.007
Carbon dioxide	0.8459
Ammonia (NH_3)	2.160

Equations of state

The thermal behavior of a gas is described by way of the correlation of pressure p, temperature T and volume V, i.e. by way of a function $F(p, V, T) = 0$. This function is called the thermal equation of state.

As well as describing the thermal behavior, a further equation, which establishes a correlation between the internal energy U and the thermal state variables, is needed to describe the caloric behavior. The caloric equation of state thus reads $U = U(V, T)$. This correlation is required when for example from the first law the resulting temperature change is to be calculated after a certain quantity of heat has been added.

Ideal gas
The thermal equation of state for an ideal gas reads:

$$pV = mRT \quad \text{or specifically}$$
$$pv = RT,$$

with the (specific) gas constant R ($R = R_m/M$) and the mass m. In addition, the following applies:

$$u = u(T) \text{ and } h = h(T).$$

In other words, the internal energy u and the enthalpy h of an ideal gas only depend on the temperature T. From the differentials of the internal energy and the enthalpy it follows with $(du/dT)_v = c_v$ and $(dh/dT)_p = c_p$:

$$u(T) = \int_{T_0}^{T} c_v(T)\,dt + u_0,$$

$$h(T) = \int_{T_0}^{T} c_p(T)\,dt + h_0.$$

This correlation can be further simplified when c_v and c_p are assumed to be constant, i.e. independent of the temperature. Such a gas is called a perfect gas. Ideal gases are governed by the correlation

$$c_p - c_v = R.$$

Table 2 shows by way of example values for c_p of some gases.

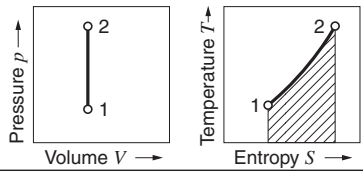
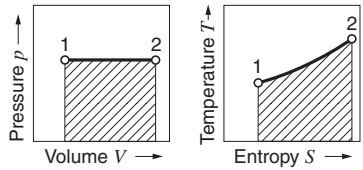
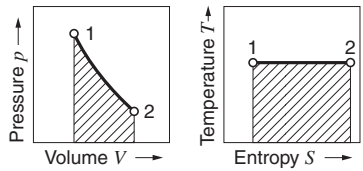
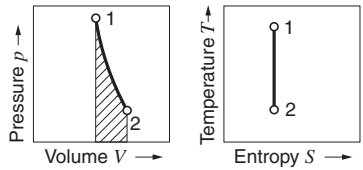
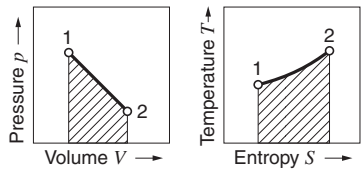
Real gas

The ideal gas equation demonstrates in certain ranges, e.g. at very high pressures, significant deviations from reality. An equation which better describes this behavior is the van der Waals equation:

$$\left(p + \frac{a}{v^2}\right)(v - b) = R\,T.$$

Here a and b are substance-specific variables, taking into account the influence of internal pressure a/v^2, resulting from molecular attraction, and the specific volume b of the molecules. There are a large number of other real gas equations as well as this equation. For these and further details, refer to the further literature [2], [3], [4].

Table 3: The most important changes of state of ideal gases and their properties

Change of state	Diagram	Equations for W_{12} volume-change work and Q_{12} added heat. (For quantities used, see Table 1).
Isochoric (V = const)		$W_{12} = 0$ $Q_{12} = m\,c_v\,(T_2 - T_1)$
Isobaric (p = const)		$W_{12} = -p\,(V_2 - V_1)$ $Q_{12} = m\,c_p\,(T_2 - T_1)$
Isothermal (T = const)		$W_{12} = -p_1 V_1 \ln \frac{p_1}{p_2}$ $Q_{12} = -W_{12}$
Reversible adiabatic ($dS = 0$)		$W_{12} = \frac{p_1 V_1}{\kappa - 1}\left(\left(\frac{V_1}{V_2}\right)^{\kappa-1} - 1\right)$ $Q_{12} = 0$
Polytropic curve ($p\,v^n$ = const)	 SAN0196-2E	$W_{12} = \frac{p_1 V_1}{n - 1}\left(\left(\frac{V_1}{V_2}\right)^{n-1} - 1\right)$ $Q_{12} = m\,c_v\,\frac{n - \kappa}{n - 1}\,(T_2 - T_1)$

Changes of state

Real changes of state of gases can be can be approached for many technical processes by simplifying, idealized assumptions. The technically most important changes of state include the isothermic (T = const), the isobaric (p = const), the isochoric (V = const), and the reversible adiabatic ($\dot{q}$ = 0 or dS = 0, no heat exchange). In addition to description by these simplified idealized changes of state, which are often realized to a good degree of approximation, it is possible, assuming an ideal gas, to describe many technical processes with greater accuracy by means of a polytropic change of state of the form

$$p\,V^n = \text{const}\,.$$

Depending on the choice of exponent n, different process paths are reproduced, thus the mentioned isothermic (n = 1), isobaric (n = 0), isochoric ($n \rightarrow \infty$), and reversible adiabatic ($n = \kappa$, with the isentropic exponent κ) changes of state. The changes of state, the associated correlations of the variables p, V and T, the associated equations for volume-change work

$$W_{12} = -\int_{V_1}^{V_2} p\,dV$$

and the added or rejected heat Q_{12} are set out in Table 3. Work and heat can further be read from the illustrated pV and TS diagrams as areas.

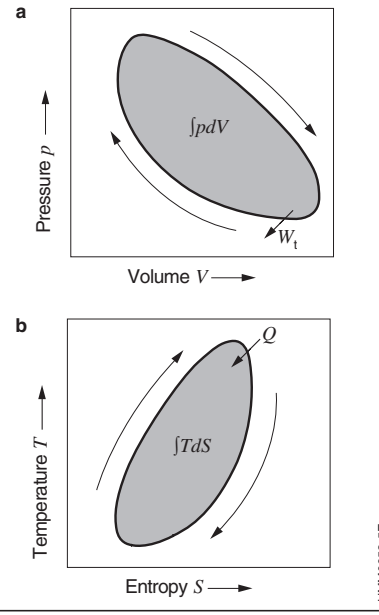
Cycles and technical applications

Basic principles

A cycle is a sequence of different thermodynamic changes of state (processes) which when completed sees the return of the initial state, i.e. the state variables in the initial and final states are identical. Here, a cycle can take place with a constant, circulating mass flow in a closed system (e.g. in a Stirling engine or a refrigerating machine) or in an open system with media exchange (e.g. in an internal-combustion engine or a gas turbine).

A cycle is typically such that either it is deprived of work (as in a heat engine) or the heat is increased by the addition of work to a higher temperature level (as in a refrigerating machine or a heat pump). These two forms can be subdivided with regard to the successive direction of their changes of state in the pV and TS diagrams into clockwise (engines) and counterclockwise machines (machines). The

Figure 2: Clockwise cycle
a) Depiction in the pV diagram,
b) Depiction in the TS diagram.



areas enclosed by the cycle represent here the input work in the pV diagram and the added heat in the TS diagram (Figure 2).

Therefore heat is added and work is output net for the situation of a clockwise cycle shown in Figure 2. The efficiency η of such a process is usually represented as the ratio of benefit (work output W_t) and expenditure (heat addition Q_{ad}), in this case therefore:

$$\eta = \frac{|W_t|}{|Q_{ad}|} .$$

For technical applications the real sequences are approached by comparison cycles. Here the entire cycle is subdivided into different reversibly accepted subcycles. The latter can be described by means of simple changes of state (e.g. isentropic, isothermal, isobaric, isochoric). In additional to simple describability, such idealized comparison cycles can be used as reference for comparing real machines as they represent the optimum in the respective process control.

Carnot cycle

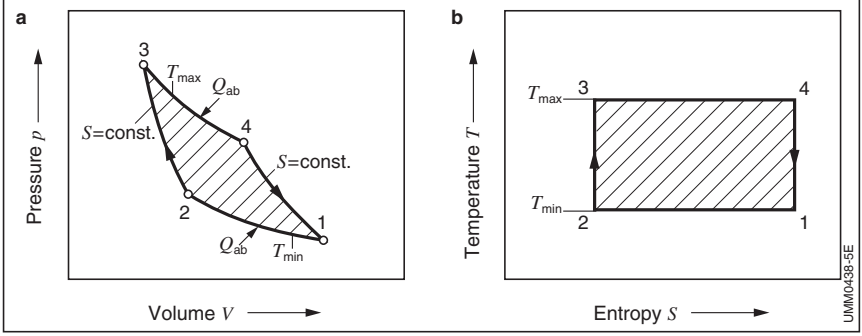
The Carnot cycle constitutes the ideal comparison cycle which shows the best possible efficiency of a machine operating between two temperature levels (T_{min} and T_{max}). Heat addition and heat rejection occur in the Carnot cycle isothermally as shown in Figure 3 and thus without dissipation, work is done (compression and expansion) in reversible adiabatic fashion.

For Carnot efficiency the following equation is obtained for an ideal gas according to the above definition and by applying the first law:

$$\eta = \frac{|W_t|}{|Q_{ad}|} = 1 - \frac{T_{min}}{T_{max}} .$$

The Carnot cycle has extra significance because – although in reality it can only be approximately represented at great expense – it describes the describes the maximum amount of utilizable energy (exergy) or non-utilizable energy (anergy) and the maximum possible efficiency can be achieved with it.

Figure 3: Changes of state of the Carnot cycle
a) In the pV diagram, b) In the TS diagram.



Comparison cycles of technical applications

Seiliger cycle

The Seiliger cycle describes the sequences and processes in internal-combustion engines. It also depicts the Diesel cycle (constant-pressure cycle) and the Otto cycle (constant-volume cycle) as boundary cycles. In the Seiliger cycle heat rejection takes place through combustion. This can be taken as isochoric (with a virtually stationary piston and constant volume) or isobaric (with a moving piston and constant in-cylinder pressure). The two forms and mixed forms of these two forms are depicted in the Seiliger cycle. The following five cycles steps are featured and depicted in Figure 4 in the pV and in the TS diagrams:

– Reversible adiabatic compression (1→2),
– Isochoric heat addition (2→3),
– Isobaric heat addition (3→4),
– Reversible adiabatic expansion (4→5),
– Isochoric heat rejection (5→1).

Important characteristics of the Seiliger cycle are the compression ratio $\varepsilon = V_1/V_2$, the pressure-increase ratio $\psi = p_1/p_2$, and the injection ratio $\varphi = V_4/V_3$. For the thermal efficiency of the Seiliger cycle the following expression is obtained for an ideal gas:

$$\eta_{th} = 1 - \varepsilon^{1-\kappa} \frac{\psi\,\varphi^{\kappa} - 1}{\psi - 1 + \kappa\,\psi(\varphi - 1)}.$$

Figure 5 shows the efficiencies that can be achieved of different process controls as a function of compression and pressure-increase ratios. It also shows the values of the boundary cases of constant-volume and constant-pressure cycles. These two

Figure 5: Efficiencies of the comparison cycles as a function of compression ratio
1 Constant-volume cycle,
2 Seiliger limit-pressure cycle, $\varphi = 1.5$, $\psi = 5.0$,
3 Seiliger limit-pressure cycle, $\varphi = 1.5$, $\psi = 1.5$,
4 Constant-pressure cycle, $\varphi = 1.5$,
5 Seiliger limit-pressure cycle, $\varphi = 2.0$, $\psi = 5.0$,
6 Seiliger limit-pressure cycle, $\varphi = 2.0$, $\psi = 1.5$,
7 Constant-pressure cycle, $\varphi = 2.0$,
ε Compression ratio,
φ Injection ratio,
ψ Pressure-increase ratio.

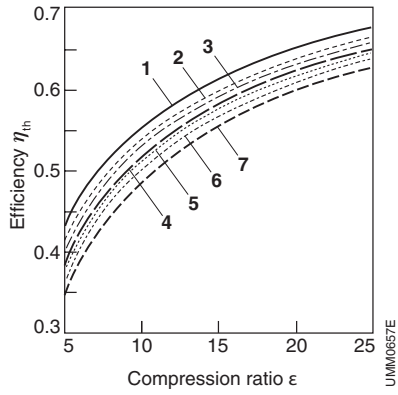
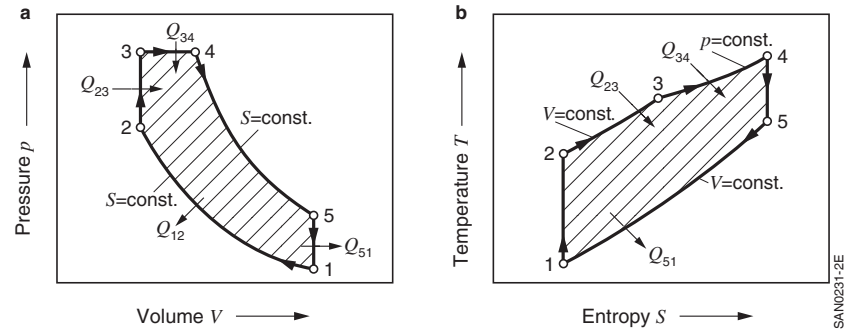


Figure 4: Changes of state of the Seiliger cycle
a) In the pV diagram, b) In the TS diagram.



boundary cases are described in further detail in the following.

Otto cycle
The Otto cycle (constant-volume cycle) represents the boundary process of the Seiliger cycle, in which the addition of heat or combustion takes place exclusively isochorically. Compared with the Seiliger cycle, therefore, the cycle step of isobaric heat addition (3→4) is absent. Process control in the pV and TS diagrams is illustrated in Figure 6. The following simplified description is obtained for the efficiency of the constant-volume cycle:

$$\eta_{th} = 1 - \varepsilon^{1-\kappa}.$$

The efficiency increases as the compression ratio ε increases. In gasoline engines the compression ratio is limited by knocking. Naturally aspirated gasoline engines operate at compression ratios of $\varepsilon = 10-12$, the maximum pressures are approximately 60 bar. Turbocharged engines, owing to the knock limit, operate at lower compression ratios, e.g. at $\varepsilon = 9-10$. Peak pressures of up to 120 bar are reached. Direct-injection stratified-charge engines have compression ratios $\varepsilon > 12$; at part load the potential for variable compression ratios is even up to $\varepsilon = 14$.

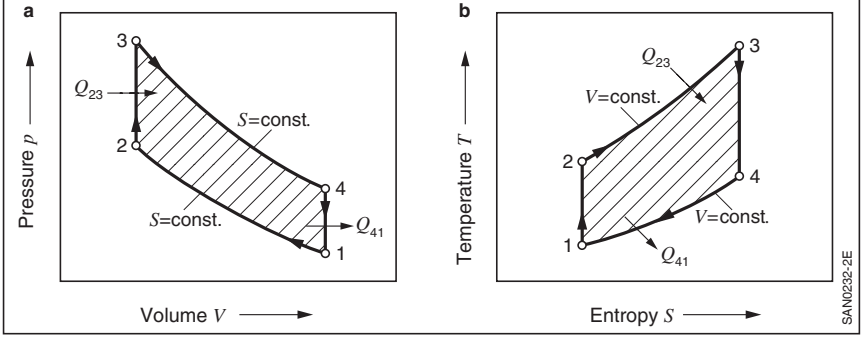
Diesel cycle
The second boundary process of the Seiliger cycle is the Diesel cycle (constant-pressure cycle). Process control in the pV and TS diagrams is illustrated in Figure 7. In this cycle the addition of heat or combustion takes place isobarically, i.e. the cycle step of isochoric heat addition (2→3) is absent. The efficiency of the constant-pressure cycle is obtained as:

$$\eta_{th} = 1 - \frac{\varphi^{\kappa} - 1}{\varepsilon^{\kappa-1}\kappa\,(\varphi - 1)}.$$

The achieved efficiency is, for the same compression ratio, lower than in the constant-volume cycle ε. However, because diesel engines are typically operated with a higher compression ratio, their efficiency is generally better. The aim of development for diesel engines is therefore to deliver a high peak pressure.

The maximum permissible peak pressures for passenger-car engines are roughly 180 bar, and for commercial-vehicle engines over 220 bar. The compression ratios for the direct-injection processes currently used today are ε =16 – 19. In the case of retarded start of injection, the final compression pressure actually corresponds to the peak pressure, and the diesel cycle resembles the constant-pressure cycle with $\varphi \approx 9$.

Figure 6: Changes of state of the Otto cycle
a) In the pV diagram, b) In the TS diagram.



Joule cycle
The Joule cycle (or Brayton cycle) is the comparison cycle for gas turbines. In contrast to internal-combustion engines and their comparison cycles, a gas turbine is subject to continuous flow. The Joule cycle is characterized by the following cycle steps (Figure 8):
– Reversible adiabatic compression (1→2),
– Isobaric heat addition (2→3),
– Reversible adiabatic expansion (3→4),
– Isobaric heat rejection (4→1).

An important characteristic of the Joule cycles is the pressure ratio $\pi = p_1/p_2$. The efficiency is thus:

$$\eta_{th} = 1 - \frac{T_1}{T_2} = 1 - \pi^{\frac{\kappa-1}{\kappa}}.$$

Clausius-Rankine cycle
The Clausius-Rankine cycle is usually used as the comparison cycle for steam turbines. This is a closed cycle in which the working medium undergoes among others two phase changes (evaporation and condensation). The cycle typically contains the following steps (Figure 9):
– Reversible adiabatic pressure increase in the liquid phase (1→2),
– Isobaric heat addition (2→3),
– Isobaric and isothermal evaporation (3→4),
– Isobaric overheating (4→5),
– Reversible adiabatic expansion (5→6),
– Isobaric and isothermal condensation (6→1).

Figure 7: Changes of state of the Diesel cycle
a) In the pV diagram, b) In the TS diagram.

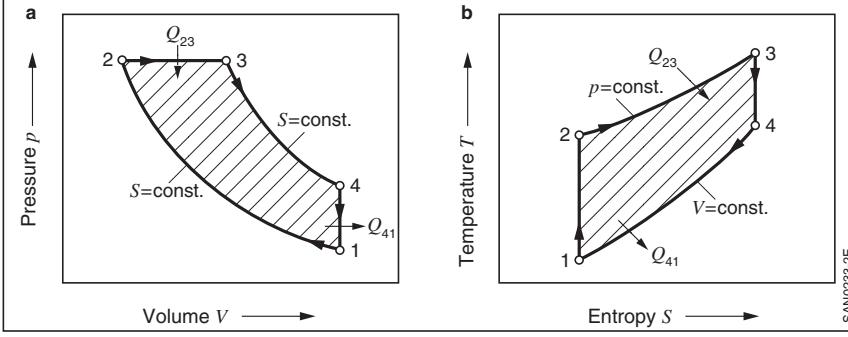
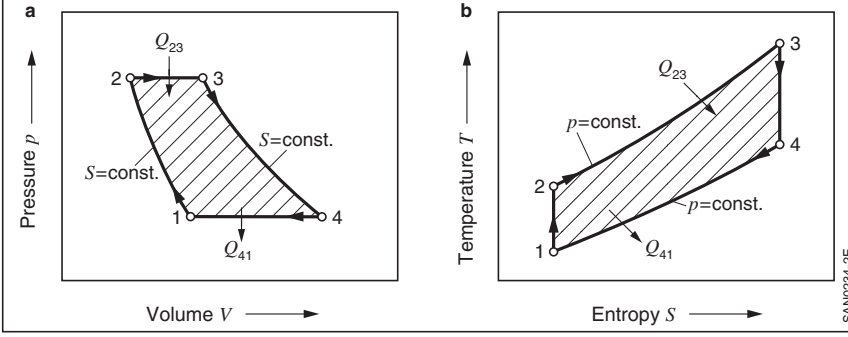


Figure 8: Changes of state of the Joule cycle
a) In the pV diagram, b) In the TS diagram.



The two curves in Figure 9 relate to the phase boundary lines (saturated-liquid and saturated-vapor lines). The saturated-liquid line represents the range of the just boiling liquid, the saturated-vapor line that of the saturated vapor. Below these two lines is the wet-steam region, i.e., the region in which liquid and gaseous phases exist together. The summit of the curves is called the critical point.

The Clausius-Rankine cycles on account of the phase changes cannot be depicted with an ideal gas; it requires a real-gas description. For the same reason, this produces a depiction which describes benefit and expenditure and thus the thermal efficiency with enthalpies. The thermal efficiency is thus [3]:

$$\eta_{th} = 1 - \frac{h_6 - h_1}{h_5 - h_2}.$$

In automotive applications the Clausius-Rankine cycle is used for example in the field of waste-heat recovery as a comparison cycle. The engine or exhaust-gas waste heat supplies the required heat here for the cycle steps 2 → 5. A turbine or a piston engine for example which makes energy available in mechanical form can be used as the engine for expansion (5 → 6).

Heat transfer

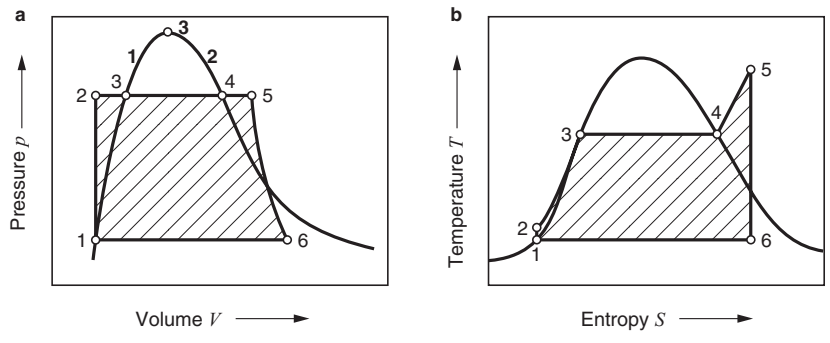
Concept formation
Essentially, heat is transferred in three different ways.

Thermal conduction
In the case of thermal conduction, heat is transferred on the molecular level. Heat is transferred here as a result of an impressed temperature gradient by a solid body or a static medium (liquid or gaseous) without macroscopic matter transport. According to the second law of thermodynamics heat is transferred here in the direction of the lower temperature.

Convective heat transfer
In the case of convective heat transfer, a flow of matter, as exists only in fluid, flowing media (liquids or gases), is required. A further distinction is made between free (caused by natural buoyancy on account of the prevailing temperature gradients) and forced (caused by externally impressed flows) convection.

Thermal radiation
In the case of thermal radiation, energy is transferred in the form of electromagnetic waves. The mechanism of heat transfer from one body to another is not matter- or substance-related and is therefore also possible in a vacuum.

Figure 9: Changes of state of the Clausius-Rankine cycle
a) In the pV diagram, b) In the TS diagram.
1 Saturated-liquid line, 2 Saturated-vapor line, 3 Critical point.



Thermal conduction

Stationary one-dimensional thermal conduction in a simple flat wall

In a homogeneous body with constant cross-sectional area A the heat flow density $\dot{q} = \dot{Q}/A$ in the one-dimensional case is given by:

$$\dot{q} = -\lambda \frac{dT}{dx}.$$

This is the Fourier thermal-conduction equation with the (substance-specific) thermal conductivity λ of the wall. For the case shown in Figure 10 (expansion in the y- and z-directions very high) it follows for a temperature difference $T_1 - T_2$ between the walls:

$$\dot{q} = \frac{\lambda}{d}(T_1 - T_2)$$

and

$$\dot{Q} = \frac{\lambda}{d}(T_1 - T_2)A.$$

By analogy to the science of electricity, the thermal-conduction resistance can be introduced:

$$R_\lambda = \frac{T_1 - T_2}{\dot{Q}} = \frac{d}{\lambda A}.$$

Table 4 shows the values for λ for some materials.

Stationary one-dimensional thermal conduction in a layered flat wall

If the wall is made not of one but of several layered materials of different thicknesses $d_1,\ldots, d_n$ and thermal conductivities $\lambda_1,\ldots, \lambda_n$ (Figure 11), this produces for a temperature difference $T_1 - T_2$ normally for the layering:

$$\dot{q} = \frac{\lambda_R}{d}(T_1 - T_2),$$

with the effective resulting thermal conductivity:

$$\lambda_R = \frac{1}{\frac{d_1}{\lambda_1} + \frac{d_2}{\lambda_2} + \ldots + \frac{d_n}{\lambda_n}}$$

and the thermal resistance:

$$R_{\lambda R} = R_{\lambda 1} + R_{\lambda 2} + \ldots R_{\lambda n}.$$

Table 4: Thermal conductivity λ of some materials at 293.15 K

Material/medium	λ [W/K·m]
Aluminum (Al)	237
Iron (Fe)	81
Copper (Cu)	399
Titanium (Ti)	22
CrNi steel (X12CrNi 18.8)	15
Glass	0.87 – 1.40
Teflon (PTFE)	0.23
Polyvinyl chloride (PVC)	0.15

Figure 10: Stationary one-dimensional thermal conduction through a flat wall

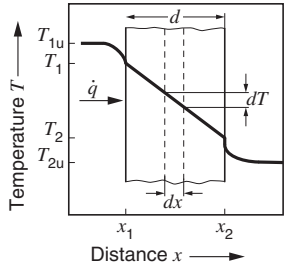
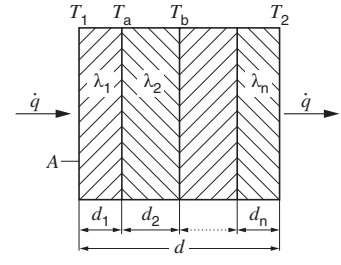


Figure 11: Stationary one-dimensional thermal conduction through a layered, flat wall



The thermal resistance therefore behaves additively (as in a series connection), with the order of the layers not playing a role.

If the temperature gradient is impressed along the plates (along a length l), it follows that:

$$\dot{q} = \frac{\lambda_p}{l}(T_1 - T_2)$$

with the effective resulting thermal conductivity

$$\lambda_p = \frac{1}{d}(\lambda_1 d_1 + \lambda_2 d_2 + \ldots + \lambda_n d_n)$$

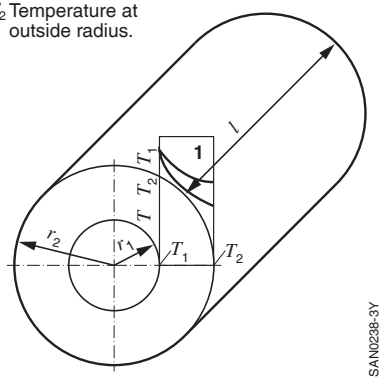
and the thermal-conduction resistance

$$\frac{1}{R_{\lambda p}} = \frac{1}{R_{\lambda 1}} + \frac{1}{R_{\lambda 2}} + \ldots + \frac{1}{R_{\lambda n}}.$$

The reciprocal value of the total thermal resistance is therefore made up additively of the reciprocal values of the individual thermal resistances and behaves as in a parallel connection.

Figure 12: Stationary one-dimensional thermal conduction through a cylindrical tube wall

1 Temperature characteristic over the tube cross-section,
r_1 Inside radius of tube,
r_2 Outside radius of tube,
l Length of tube,
T_1 Temperature at inside radius,
T_2 Temperature at outside radius.



Stationary one-dimensional thermal conduction in a tube wall

Similarly to thermal conduction through a wall, the following is obtained for a tube (of virtually infinite length) of length l in the one-dimensional case (thermal conduction only normally to the tube axis) as shown in Figure 12:

$$\dot{Q} = 2\pi l \lambda \frac{T_1 - T_2}{\ln\left(\frac{r_2}{r_1}\right)},$$

where T_1 is the temperature on the inner side and T_2 the temperature on the outer side.

Because of the different surface areas on the inner and outer sides of the tube the temperature assumes a logarithmic characteristic in the tube wall.

Stationary heat transmission

In many technical applications thermal conduction inside a solid body occurs in combination with a convective heat transfer (see Convective heat transfer). A (convective) heat transfer occurs as well as the thermal conduction, as shown in Figure 10; towards the wall on account of the temperature difference $(T_{1u} - T_1)$ and away from the wall on account of the temperature difference $(T_2 - T_{2u})$. The heat flows must naturally all be of equal magnitude, since heat is not lost in either convective heat transfer or in thermal conduction. The heat flows that occur are governed by:

$$\dot{Q} = \frac{\lambda}{d}(T_1 - T_2)A,$$

$$\dot{Q} = \alpha_{1u}(T_{1u} - T_1)A,$$

$$\dot{Q} = \alpha_{2u}(T_2 - T_{2u})A,$$

with the heat-transfer coefficients α_{1u} and α_{2u}. For such a situation it is possible to create the heat-transmission coefficient k, which is obtained as follows:

$$\frac{1}{k} = \frac{d}{\lambda} + \frac{1}{\alpha_{1u}} + \frac{1}{\alpha_{2u}}.$$

While taking into account the two fluid temperatures T_{1u} and T_{2u}, the following is obtained for the heat flow:

$$\dot{Q} = k(T_{1u} - T_{2u})A.$$

Convective heat transfer
The heat-transfer coefficient α has already been introduced. It is dependent on a variety of influencing variables, such as, for example, temperature, density, geometry, type of flow (laminar or turbulent), and speed. It is determined by the conservation equation (nonlinear partial differential equations) describing the (fluidic) problem for which there are only closed analytical solutions in exceptional cases. Heat-transfer coefficients are therefore often determined on the basis of numerical solutions or directly from experiments. The functional dependencies for the heat-transfer coefficients can be found from these on the basis of dimensional analysis and similarity relationships.

In view of the complexity, reference is made at this point for derivation and further details to the relevant literature (e.g. [1]). The most important dimensionless parameter for heat transfer is the Nusselt number Nu:

$$Nu = \frac{\alpha L}{\lambda}.$$

This numbers contains, in addition to the heat-transfer coefficient α, the thermal conductivity λ of the fluid, and a reference length L, which can be for example a tube diameter or a plate length. It can further be shown that the following functional correlations apply to the averaged, i.e. integrated over the body surface and thus location-independent Nusselt number Nu_∞:
– $Nu_\infty = f(Re, Pr)$ for forced convection,
– $Nu_\infty = f(Gr, Pr)$ for free convection.

The Nusselt number and thus the heat transfer can therefore be described as a function of the dimensionless parameters Reynolds number Re, Prandtl number Pr and Grashof number Gr. These must be calculated in each case as a function of the geometry and the flow (Re and Gr) or as a function of the fluid properties (Pr). The Reynolds number is defined as:

$$Re = \frac{cL}{\nu}.$$

Here c is the flow velocity, L a length characteristic of the application (e.g. tube diameter), and ν the kinematic viscosity of the fluid.

The Prandtl number is a pure substance variable and is given by:

$$Pr = \frac{\nu}{a},$$

with the thermal diffusivity a. For gases Pr = 0.7 can be assumed by approximation for pressures below 10 bar.

To calculate the Grashof number Gr, reference is made to the relevant literature (e.g. [1]).

The correlations for some selected cases with forced convection are set out in the following. The correlations in each case for the averaged values together with their range of application are specified here. The heat-transfer coefficients can be calculated with the Nusselt numbers determined from them and the definition of these.

Flat plate with laminar boundary layer subject to longitudinal flow
Nusselt correlation:

$$Nu_\infty = 0.664\, Re^{\frac{1}{2}} Pr^{\frac{1}{3}}.$$

Characteristic length:
Plate length L.

Range of application:
$Re \le 5 \cdot 10^5$; $0.6 \le Pr \le 2{,}000$.

Flat plate with turbulent boundary layer subject to longitudinal flow
Nusselt correlation:

$$Nu_\infty = 0.037\, Re^{0.8} Pr^{\frac{1}{3}}.$$

Characteristic length:
Plate length L.

Range of application:
$5 \cdot 10^5 \le Re \le 10^7$; $0.6 \le Pr \le 60$.

Turbulent pipe inner flow
Nusselt correlation:

$$Nu_\infty = 0.023\,Re^{0.8}\,Pr^n,$$

with $n = 0.4$ for $T_W > T_F$
and $n = 0.3$ for $T_W < T_F$.

Here, T_W is the wall temperature and T_F the averaged fluid temperature. The Nusselt number must be evaluated for the averaged fluid temperature because this changes during an inner flow from inlet to outlet.

Characteristic length:
Tube diameter D.

Range of application:
$10^4 \le Re$; $0.7 \le Pr \le 120$.

The tube length must be at least ten times the tube diameter.

The above correlation can also be used for tubes with non-circular cross-sections; the equivalent or hydraulic diameter d_h, which is created as follows, is used here:

$$d_h = \frac{4A}{U},$$

with the through-flow cross-sectional area A and the circumference U.

Thermal radiation
Thermal radiation depends only on the type and the temperature of the radiating body. When radiation strikes a body, the following phenomena are observed: Part of the incident radiation is reflected (reflection), part is absorbed (absorption), and part is let through (transmission). Due to the conservation of energy, it follows that the sum of these three energy components is equal to the amount of energy of the incident radiation.

According to the Stefan-Boltzmann law, the heat flow radiated from a body with surface A at temperature T is:

$$\dot{Q} = \varepsilon \sigma A T^4.$$

with the Stefan-Boltzmann constant σ,

$$\sigma \approx 5.67 \cdot 10^{-8} \frac{\mathrm{W}}{\mathrm{m^2K^4}},$$

and the emissivity ε of the body. The emissivity is between 0 (total reflection) and 1 (black-body radiator) and depends among others things on the temperature and the surface condition of the body. Some emissivity values are given by way of example in Table 5.

References
[1] H.D. Baehr, K. Stephan: Wärme- und Stoffübertragung. Springer, 9th Edition, Verlag Springer Vieweg, 2016.
[2] H.D. Baehr, S. Kabelac: Thermodynamik: Grundlagen und technische Anwendungen. 16th Edition, Verlag Springer Vieweg, 2016.
[3] E. Hahne: Technische Thermodynamik: Einführung und Anwendung. 5th Edition, Oldenbourg Wissenschaftsverlag, 2010.
[4] B. Weigand, J. Köhler, J. von Wolfersdorf: Thermodynamik kompakt. 4th Edition, Springer Vieweg, 2016.

Table 5: Emissivity ε of some materials (values in the range up to 300 °C)

Black-body radiator	1.00
Aluminum, unmachined	0.07
Aluminum, polished	0.04
Cast iron, rough, oxidized	0.94
Cast iron, turned	0.44
Copper, oxidized	0.64
Copper, polished	0.05
Brass, matt	0.22
Brass, polished	0.05
Steel, matt, oxidized	0.96
Steel, polished, oil-free	0.06
Steel, polished, oiled	0.40

Electrical engineering

The history of electricity goes back around 3,000 years. For a long time electrical effects were subject to mysterious interpretations. Only in the 19th and 20th centuries were scientists able to describe these effects more accurately through scheduled experimentation and exact measurement. It was recognized that most processes run according to certain laws. The sensible application of these laws resulted in the rapid development of electrical engineering and – for some decades now – electronics. Electrical engineering predominantly comprises areas in which the energy content of electricity plays a role (e.g., in electric motors) while electronic devices are used to measure and control this energy. There are numerous uses for devices which comprise electrical and electronic components. A clear distinction between these two areas is often not possible.

Table 1: Quantities and units
(Additional quantities and units in the text)

Quantity		SI unit
A	Area	m^2
a	Distance	m
B	Magnetic flux density, induction	$T = Wb/m^2 = V\,s/m^2$
C	Electrical capacitance	$F = C/V$
D	Electric flux density, electric displacement	C/m^2
E	Electric field strength	V/m
F	Force	N
f	Frequency	Hz
G	Electrical conductance	$S = 1/\Omega$
G	Antenna gain	dB
H	Magnetic field strength	A/m
I	Electric current	A
J	Magnetic polarization	T
k	Electrochemical equivalent	kg/C (usually: g/C)
L	Inductance	$H = Wb/A = V\,s/A$
l	Length	m
M	Electric polarization	C/m^2
P	Power	$W = V\,A$
P_s	Apparent power	V A
P_q	Reactive power	var = V A
Q	Quantity of electricity, electric charge	$C = A\,s$
q	Cross-sectional area	m^2
R	Electrical resistance	$\Omega = V/A$
r	Radius	m
r	Reflectance factor	–
S	Electromagnetic power density	W/m^2
s	Standing wave ratio	–
T	Temperature	K
t	Time	s
U	Voltage	V
V	Magnetic potential difference	A
W	Work, energy	$J = W\,s$
W_e	Electrical energy	Ws
W_m	Magnetic energy	Ws
w_e	Electrical energy density	Ws/m^3
w_m	Magnetic energy density	Ws/m^3
w	Number of turns	–
Z	Characteristic impedance	Ω
α	Geometric angle	° (degrees)
ε	Dielectric constant	$F/m = C/(V\,m)$
ε_0	Electric field constant $\approx 8.854 \cdot 10^{-12}$ F/m	
ε_r	Relative permittivity	–
λ	Wavelength	m
Θ	Current linkage	A
μ	Permeability	$H/m = V\,s/(A\,m)$
μ_0	Magnetic field constant $\approx 1.257 \cdot 10^{-6}$ H/m	
μ_r	Relative permeability	–
ρ	Resistivity	$\Omega\,m = 10^6\,\Omega\,mm^2/m$
σ	Electrical conductivity (= $1/\rho$)	$1/(\Omega\,m)$ $= 10^{-6}\ m/(\Omega\,mm^2)$
Φ	Magnetic flux	$Wb = V\,s$
φ	Phase displacement angle	° (degrees)
φ (P)	Potential at point P	V
ω	Angular frequency (= $2\,\pi\,f$)	Hz

Basic principles of electricity

Atomic structure

All matter consists of tiny building blocks called atoms (atomos, Greek: indivisible). A copper cube with an edge length of 1 cm contains approximately 10^{23} atoms which are permanently "bound" to each other. According to the findings of atom physics atoms are made up of even smaller particles – electrons, protons, and neutrons. Our solar system, in which the planets (Earth, Jupiter, Mars, etc.) orbit a common nucleus, the Sun, serve as the model for the atomic structure. Materials which are composed of atoms of the same type are called elements.

Today we know that electricity is a property of electrons. These smallest particles orbit the nucleus, consisting of protons and neutrons, at high speed. Depending on the element in question, its atoms contain a different number of electrons on different paths, the electron shells. An electron carries the smallest occurring quantity of electricity; the designation "negative" was coined for this type of electric charge.

The nuclear particles are neutrons and protons. The latter carry the same quantity of electricity as the electrons. But because their electric charge is reversed, they behave like electrons and are positive. These positive protons determine, together with the electrically neutral neutrons, the quantity of electricity.

Because the electrons orbit the nucleus at high speed, they would be thrown from their path by centrifugal force if the two opposing charges in the atom (electrons negative, protons positive) were not mutually attracted. Conversely, bodies with the same electric charge repel each other.

Charge carriers

Free electrons

The movement of these conduction electrons can be guided by electrical forces (voltage) in a certain direction to obtain the electron flow. By supplying energy (electrical, thermal, optical, etc.) it is possible even to release electrons under certain preconditions from the atomic union of a material.

The electrons on the outermost shell of the atoms determine both the electrical and the chemical behavior of a material and establish the bond with neighboring atoms of other or identical elements. These electrons participating in the bonds are called valence electrons. Because only valence electrons participate in electrical processes, the atom model can be further simplified. Accordingly, this model consists of the electrically effective electrons of the outermost shell and a positive kernel, which contains the nucleus and the electrons of the remaining shells. In the copper model each kernel has a valence electron.

Ions

Special measures, e.g., mechanical or chemical effects, can be used to separate or add one or more valence electrons in an atom. This atom is then no longer electrically neutral and is known as an ion. Atoms that lack electrons (negative charge) are positive ions while those that have more electrons than protons are negative ions.

Missing electrons in the atomic union

Finally, a third kind of electric charge carrier takes effect in semiconductors (diodes, transistors). Points in the atomic union with missing electrons (holes) acts as positive charge carriers. Unlike ions, however, the kernels in semiconductor materials (germanium, silicon) are not freely moving.

Charge quantity
The smallest possible charge quantity is the elementary charge; the electron (or the proton) is the carrier of this elementary charge. Each charge quantity is a whole multiple of the elementary charge and the (letter) symbol for the charge quantity is Q.

In contrast to water droplets, it is not the mass of the electrons that is considered here, but only its electrical property, the charge. This has a quantitative character and quantity of electricity is therefore synonymous with electric charge. The unit of charge quantity is the "coulomb" with the symbol C.

Another term for "coulomb" is the "ampere-second" with the symbol "As".

1 C = 1 As; $[Q]$ = C = As.

The amount of the elementary charge is:

$e = 1.602\,176\,634 \cdot 10^{-19}$ As.

This value is obtained from the already established unit "ampere" in accordance with the SI system (see Quantities and units).

For the most part billions of tiny charge carriers (electrons) participate in electrical and electronic processes. It is therefore natural to distill a large number of elementary charges down to a "handy" measure. From the size of elementary charge e: 6.3 trillion elementary charges equal 1 coulomb.

Electrical quantities in an electric circuit

Current
Current intensity
Electric current is the transfer of charge. "Transferred charge quantity" refers to the charge quantity Q or ΔQ that flows through a conductor cross-section A over time t or Δt. The current intensity I is defined as the transferred charge quantity per time unit ([1], [2]):

$$I = \frac{Q}{t}$$

Unit of current intensity:

$[I]$ = A.

Thus for the unit of charge quantity:

$[Q] = [I] \cdot [t]$ = As = C.

A voltage source and a complete circuit are required to generate a current flow (Figure 1). Only negative charges (electrons) are transferred. The current intensity in a complete circuit (without junctions) is the same at all points.

The current flow can be discerned when an electric light bulb lights up. All electrical effects involve the progression of electric charge carriers. Negative and positive charge carriers often participate in the conduction processes in the same device. Normally only the conduction electrons in metals are considered.

Technical current direction
That direction in which positive charge carriers were to move it they were present is defined as the technically positive current direction. The technical current direction is therefore opposite to the direction of electron movement (Figure 1). In the case of the movement of positive ions, the technical current direction and the direction of movement of the charge carriers are identical.

Current density
The current density S is the current intensity I referred to the conductor cross-section A. Figure 2 shows the current density S in the electric conductor.

Current density $S = \frac{I}{A}$.

S Current density in A/mm^2,
A Conductor cross-section in mm^2.

Voltage
Voltage is obtained when the positive charge carriers and the negative charge carriers (electrons) which are present in the electrically neutral atoms of all materials are separated from each other. This occurs in voltage generators, e.g., in a flashlight battery. This battery exhibits a surfeit of electrons (negative charge) at the negative terminal and a deficit of electrons (positive charge) at the positive terminal. The negatively charged electrons of a connected conductor are therefore attracted by the positive terminal and repelled by the negative terminal. Because a large number of conduction electrons is available in the conductor, this effect continues and an electron stream flows from the negative terminal to the positive terminal.

The voltage is the cause of the current. The current direction depends on the polarity of the voltage and on the type of charge carriers participating in the current flow. Positive ions in an electrically conductive liquid move for example from the positive terminal (repulsion) to the negative terminal (attraction), i.e., in the opposite direction to the electron stream in a conductor. The positive direction of the technical current is defined by a standard from positive to negative.

Crucial to the occurrence of a voltage are different electric charges at the terminals of the voltage generators. The effect of the voltage corresponds to a force that sets the electric charge carriers (e.g., conduction electrons) in motion. Therefore voltage at the point where it is generated is also called electromotive force. It is recognized that without voltage (cause) there can be no current flow (effect). In contrast, voltage can be present even if there is no current flowing, e.g., between the terminals of a voltage generator.

Figure 1: Electric circuit with generator and light bulb
G Generator (voltage source).

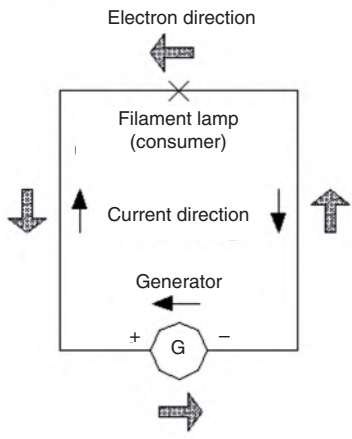
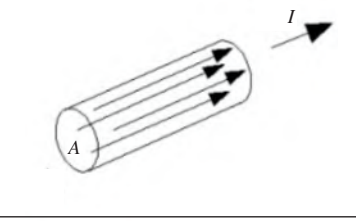


Figure 2: Current density S in an electric conductor
A Cross-sectional area of conductor,
I Electric current.



Unit of measurement
The unit of measurement for voltage, i.e., for a certain charge difference, is the volt and the measurement symbol is V. The symbol (abbreviated form for voltage) is U.

Charge transfer
Voltage is the cause of the movement of charge carriers. A copper wire with a length of 1 m and a cross-section of 1 mm^2 has a volume of 1 cm^3 and contains approximately 10^{23} conduction electrons. At a current intensity of 1.6 A around 10^{19} electrons would flow per second. This constitutes around 1/10,000 of the charge carriers present in the conductor. If now around

10^{19} electrons flow out at the conductor end in one second, the electrons must move further along in the conductor in the same period of time by around 1/10,000 of the total wire length (l = 1 m = 1,000 mm), i.e., by around 0.1 mm. This produces a speed of approximately 0.1 mm/s! At this speed an electron to move through a wire 1 m long needs around 10,000 seconds (approximately three hours!).

Practice shows that a switched-on desk lamp for example lights up instantly when it is connected to the power socket. Voltage as the cause of current must therefore cover the distance (often several meters long) from the power socket through the connecting cable to the lamp at very high speed. Because the electrons are very numerous in the conductor, they "bump" into each other when a voltage is applied. This bump effect (voltage effect) is reproduced at a speed that can approach the speed of light ($3 \cdot 10^8$ m/s).

Resistance

Line resistance

The voltage U can only then set in motion the electrons in a conductor when the circuit is complete. The current intensity here depends on the effective voltage and on the resistance offered to the electric current by the conductor. This "braking effect" varies depending on the material, length, and cross-section of a line. The unit of measurement for resistance is the ohm. The Greek letter Ω is used as the measurement symbol. The symbol is R. Resistance R is all the greater

– the greater the length l of a conductor,
– the smaller the cross-sectional area A,
– the greater the resistivity ρ, i.e., the more poorly the material in question conducts the electron stream under otherwise identical conditions.

For resistance R:

$$R = l/\gamma A,$$

$$R = l\rho/A.$$

R Line resistance in Ω,
l Conductor length in m,
A Conductor cross-section in mm²,
ρ Resistivity in (Ω mm²)/m,
γ Conductivity in m/(Ω mm²).

Resistivity

The resistivity ρ denotes the material. Table 2 specifies for some materials the resistivity ρ and its reciprocal value, the conductivity γ. Different materials must be compared with each other under identical conditions: 1 m length, 1 mm² cross-section, and 20 °C temperature are the conditions for comparing the resistivity of wires.

Conductance

Increasing the cross-section of a conductor reduces its resistance, i.e., it conducts the electron stream better. A low resistance R corresponds to a high conductance G, which is obtained as the reciprocal of the resistance. The unit of measurement for conductance is the siemens.

$$G = \frac{1}{R},$$

G Conductance in S (siemens),
R Resistance in Ω (ohm).

The electric current is conducted well in a circuit with high conductance or low resistance.

Table 2: Resistivity and conductivity of some materials

Material	Resistivity ρ [Ω mm²/m]	Conductivity γ [m/Ω mm²]	Temperature coefficient α [1/K]
Silver	0.0164	61	0.0038
Copper	0.01724	58	0.00393
Aluminum	0.0278	36	0.00403
Iron	0.13	7.7	0.0065
Constantan	0.5	2.0	±0.00001

Example: Which conductance belongs to the resistance value 2 Ω?

$$G = \frac{1}{R} = \frac{1}{2\Omega} = 0.5\ \text{S}.$$

Temperature coefficient of resistors

The resistivity ρ is specified in each case for a temperature or 20 °C. Most materials change their resistance when their temperature is altered.

When metals are heated, the atomic union becomes slightly unsettled by the energy supply. This restricts the mobility of the conduction electrons. The resistance of most metals increases as the temperature rises. Up to a temperature of around 200 °C the resistance change is proportional to the temperature change. In most liquids, gases, and semiconductors the resistance decreases as temperature increases.

The resistance change ΔR per degree temperature change is denoted by the temperature coefficient α. The unit of measurement for temperature differences between two temperature points is °C or K (kelvin) on the celsius or the kelvin scale. Percentages for the temperature coefficient in many cases make it easier to assess the resistance change as a function of the temperature change.

The temperature-dependent resistance change is governed by the following relationship:

$$\Delta R = \alpha R_{20}\, \Delta T$$

ΔR Resistance change in Ω,
ΔT Temperature change in K or C,
α Temperature coefficient in 1/K,
R_{20} Resistance at 20 °C.

Most pure metals have a positive temperature coefficient. Table 2 shows, in addition to the values for the resistivity ρ and the conductivity γ, the temperature coefficient α for some metals.

Figure 3 shows the characteristic curves of thermistors. Materials with a positive temperature coefficient conduct better at low temperatures and are therefore called PTC (positive temperature coefficient) thermistors. Materials with a negative temperature coefficient conduct better at high temperatures and are therefore called NTC (negative temperature coefficient) thermistors.

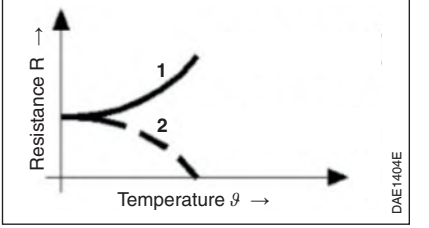
As well as the ambient temperature, the current intensity can also influence the resistance of a material. "Electric friction" between moving charge carriers and kernels creates in the current-carrying material heat and with it a temperature increase, which increases or decreases the resistance depending on the type of material.

While the resistance of pure metals decreases with decreasing temperature, the resistance of semiconductors increases. In the area surrounding absolute zero (0 K ≈ –273 °C) metals exhibit no resistance whatsoever (superconductivity) and semiconductor materials achieve a infinitely high resistance. In complicated devices temperatures in the vicinity of absolute zero have already been obtained.

Temperature coefficient of measuring resistors

The resistance of measuring resistors are only permitted to change very slightly in response to temperature fluctuations. The change has an affect on the electrical properties of measuring devices which would be influenced by temperature. Very small temperature coefficients for resistor materials are obtained by alloying certain metals, e.g., copper, nickel, and manganese. The temperature coefficient is, depending on the content of alloying constituents, between –0.001 and –0.005 % per K.

Figure 3: Characteristic curves of thermistors
1 PTC thermistor,
2 NTC thermistor.



Temperature coefficient of semiconductors
Virtually all semiconductor materials exhibit a negative temperature coefficient. This ranges between –2 and –5 % per K for germanium and silicon.

Capacitor
Design
Capacitors are components which in principle consist of two opposing, conductive plates (plate capacitor). The two plates are separated by an insulating layer (dielectric). When capacitors are used the effect of the electric field between the two capacitor plates is exploited. When a voltage is applied, charge quantities can be stored with the storage capacity of a capacitor being dependent on the following factors:
– the size and constitution of the plate surface,
– the distance between the plates, and
– the conductivity of the dielectric for the electric field lines.

Capacitance
The possibility of storing electrical energy in an electric field with a capacitor is utilized in varied ways in electric and electronic circuits. A measure of the capacity to absorb electric charges is the capacitance C; this has the unit "farad" (F).

The charge quantity Q absorbed by the capacitor is all the greater the greater its capacitance C and the higher the applied voltage U are as the causal charge difference. The charge quantity Q is calculated from

$$Q = C\,U.$$

The capacitance C can thus be specified as the ratio of the absorbed charge Q to the applied voltage U.

Calculation formula for capacitance
The capacitance C depends on the design of a capacitor, the plate area A, the distance between the plates d, the electrical properties of the dielectric, and the so-called dielectric constant ε as follows.
– Effective plate area A: The greater the plate area, the greater the capacitance. A linear relationship applies → $C \sim A$.
– Distance between the plates d: The smaller the distance between the plates, the greater the capacitance. An inversely proportional relationship applies → $C \sim 1/d$.
– Dielectric constant ε: The greater ε is ("better" dielectric), the greater the capacitance → $C \sim \varepsilon$.

The electric constant ε_0 specifies the permittivity of free space (vacuum between the capacitor plates); it has the constant value

$\varepsilon_0 \approx 8.85$ As/Vm.

The dielectric coefficient ε_r specifies the factor by which the capacitance increases for a certain dielectric in relation to vacuum. It is material-specific and has the unit 1. The dielectric constant ε is obtained as the product of these two quantities.

$\varepsilon = \varepsilon_0\,\varepsilon_r$, with the unit As/Vm or F/m.

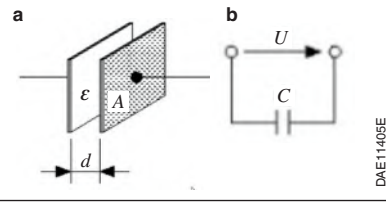
The calculation formula for the capacitance of a capacitor can therefore also be specified as follows:

$$C = \varepsilon_0\,\varepsilon_r\,A/d\ .$$

Energy content
The energy content of a charged capacitor (charge Q, voltage U, capacitance C) is:

$$W = \frac{1}{2}QU = \frac{Q^2}{2C} = \frac{1}{2}CU^2\ .$$

Figure 4: Design and graphical symbol of a capacitor
a) Design,
b) Graphical symbol with applied voltage U.



Coil

Design

A current-carrying conductor generates a magnetic field whose magnetic field lines are arranged concentrically around the conductor (see Magnetic field).

In a coil a wire is wound on a coil core in such a way as to produce many parallel conductor loops. As a result, the magnetic field caused by a current is amplified (for details, see "Coil in the alternating-current circuit").

Inductance

The inductance L is a characteristic quantity of coils. It is dependent on the geometric properties of the coils, the number of turns, and the magnetic properties of the material filled by the coil core. For a long coil of length l, cross-sectional area A, and number of turns N:

$$L = \mu_0 \mu_r N^2 A / l.$$

μ_0 is the magnetic constant with the value $4\pi \cdot 10^{-7}$ Vs/Am, μ_r is the material-dependent permeability. For air $\mu_r = 1$.

Energy content

The energy content of the coil carrying current I of inductance L is:

$$W = \frac{1}{2} L I^2.$$

Conductors, nonconductors, semiconductors

The fact that the resistance value R increases as the length increases and the cross-section decreases is common to all materials. The following may be imagined: Aside from the electric charge carriers for the most part a large number of atoms (kernels) is present. In metals the conduction electrons must really "squeeze through" between the kernels. It is true that this "electric friction" is lower in a short piece of wire with a large cross-section than in a long conductor with a small cross-section.

Elements differ in the structure of their atoms. Valence electrons establish the bond with identical or different neighboring atoms. Depending on the material, there is a different number of effective electric charge carriers (e.g., conduction electrons in metals) for a specific volume. The number of effective charge carriers per unit of volume and their mobility between the kernels of the atomic union determine the resistivity ρ of a material.

Conductors

Most metals are good conductors because they have a large number of conduction electrons per unit of volume. The resistivity of metals is therefore low. However, there are significant differences between different metals (see Table 2).

Copper is the preferred choice of conductor due to its low resistivity. Aluminum, nickel, platinum, gold, tantalum, tungsten, silver, mercury, and tin are often used as well. In certain electric motors brushes made of carbonaceous materials are required and in simple microphones sound-pressure changes are converted into resistance changes with the aid of carbon granules. A voltage drop is frequently desirable in electrical engineering and in electronics; the resistor materials are used here to reduce the current intensity. The corresponding components are pooled together under the generic term "resistors". These resistors are – depending on the requirements – made of certain metals, carbon or metal alloys.

Nonconductors

In nonconductors the valence electrons are bound by strong forces within the atomic structure and only very few free charge carriers are available, resulting in high resistivity. These bonds can however be broken by high voltages or by strong heating. Nonconductors are called insulators; these include for example porcelain, glass, rubber, plastics, Teflon, and PVC.

A distinction is made between conductors and nonconductors, depending on the number of effective electric charge carriers.

Semiconductors

There are materials whose conductivity cannot be classified with either conductors or nonconductors. They are, like the components made from them, called semiconductors. These semiconductors (germanium, silicon) are in their very pure state good insulators. The conductivity can be greatly altered by specific measures, e.g., by targeted contamination or temperature increase. The properties of semiconductor materials are used in a variety of applications in diodes and transistors.

Laws

Ohm's law

In an electric circuit current, voltage, and resistance are linked by Ohm's law. This law defines the relationship between voltage U and current I in solid and liquid conductors. Ohm's law states:

$$U = R I .$$

The proportionality constant R is called ohmic resistance, and is measured in ohms (Ω).

Ohm's law contains the quantities current intensity I, voltage U, and resistance R. One quantity can be calculated when the other two are known. Ohm's law is one of the important basic principles of an electronics engineer.

Kirchhoff's laws

1st Kirchhoff law: current law

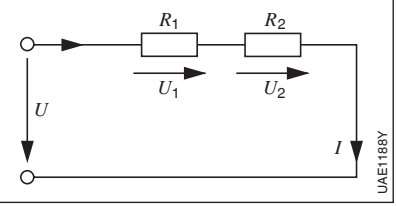
The sum of currents (according to their counting direction) flowing into each junction (node) is equal to the sum of currents flowing out of that junction.

2nd Kirchhoff law: voltage law

For each closed loop of a conductor network the sum of the component voltages oriented in the direction of the loop at the individual elements (resistors and sources) is equal to the sum of the component voltages oriented against the direction of the loop.

Figure 5: Series connection of resistors

U Voltage,
I Current,
R Resistance.



Connections of resistors

Series connection of resistors

The series connection of resistors entails the connection of a number of resistors in succession (Figure 5).

Measurements at different points in a series connection confirm that the current intensity is identical at all points. The following generally applies to series connection:

$I = I_1 = I_2 = I_3 = \ldots$; all currents in A.

In contrast to current measurements, it is not necessary in the case of voltage measurements to break the circuit. For this reason the preferred option for troubleshooting or testing (e.g., in electronic devices and systems) where possible is voltage measurement.

In the example from Figure 5 a specific partial voltage (U_1, U_2) is measured at each resistor. The sum total of these partial voltages is just as high as the total voltage U at the circuit input and the following generally applies:

$U = U_1 + U_2 + U_3 + \ldots$; voltages in V

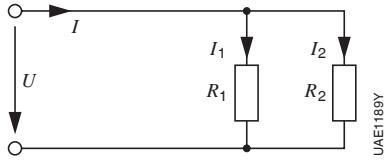
The total resistance R of this circuit results, also according to Ohm's law, from the total voltage U and the current intensity I as

$R = R_1 + R_2 + R_3 + \ldots$

Parallel connection of resistors

The parallel connection of resistors entails current division (Figure 6). According to the current law (1st Kirchhoff law) the following generally applies:

Figure 6: Parallel connection of resistors
U Voltage,
I Current,
R Resistance.



$I = I_1 + I_2 + I_3 + \ldots$;

$\frac{1}{R_{tot}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}$ or

$G_{tot} = G_1 + G_2 + G_3 + \ldots$

The same voltage U is applied at all the resistors.

Example:
Each individual resistor has 100 Ω, the total resistance for three resistors is 33.33 Ω. The values can be calculated by way of two methods, namely via the conductance in S (siemens) or via the resistance in Ω.

$G = \frac{1}{R} = 1/100\ \Omega = 0.01$ S.

$G = G_1 + G_2 + G_3$
$= 0.01$ S $+ 0.01$ S $+ 0.01$ S $= 0.03$ S.

This results in a total resistance of

$R = \frac{1}{G} = 1/0.03$ S $= 33.33\ \Omega$.

The same voltage is effective at each single resistor of a parallel connection. When different resistors are connected in parallel, different single currents also result. According to Ohm's law the current intensity is greatest in the smallest resistor. Conversely, the smallest current flows in the largest resistor. The currents are inversely proportional to the resistors.

Connections of capacitors
Parallel connection of capacitors
Parallel connection of capacitors increases the plate area A (Figure 7b) and total capacitance. The following applies:

$$C_{tot} = C_1 + C_2 + \ldots + C_n.$$

C_{tot} Total capacitance in F.
C_1, C_2 … C_n Individual capacitances in F.

Series connection of capacitors
Series connection of capacitors (Figure 7a) increases the distance between the outer linings. The total capacitance decreases as a result. The following applies:

$$\frac{1}{C_{tot}} = \frac{1}{C_1} + \frac{1}{C_2} + \ldots + \frac{1}{C_n}.$$

Connection of coils
Series connection of coils
Series connection of coils (Figure 8a) entail the adding up of the individual inductances.

$$L_{tot} = L_1 + L_2 + \ldots + L_n.$$

L_{ges} Total inductance in H.
L_1, L_2 … L_n Individual inductances in H.

Parallel connection of coils
The following applies to the parallel connection of coils (Figure 8b):

$$\frac{1}{L_{tot}} = \frac{1}{L_1} + \frac{1}{L_2} + \ldots + \frac{1}{L_n}.$$

Electric power and work

In a resistor through which current passes, the following holds for the energy converted during time t into heat or another form of energy (with ohmic resistance R, voltage U and current I):

$$W = UIt = RI^2 t$$

and thus for power:

$$P = UI = RI^2 = U^2/R.$$

A voltmeter and ammeter can be used to determine the electric power of a device with reference to these mentioned formulas (Figure 9). Figure 10 shows the circuit with a connected wattmeter.

Example
What is the power consumption of an electrical device from a voltage of U = 230 V and I = 0.5 A?

$$P = UI = 230\ \text{V} \cdot 0.5\ \text{A} = 115\ \text{W}$$

The electrical work W is the product of the electric power $U \cdot I$ and the expended time t.

$$W = P\,t.$$

W Electrical work in Ws, Wmin or Wh,
P Power in W,
t Time in s, min or h.

Figure 7: Connection of capacitors
a) Series connection,
b) Parallel connection.

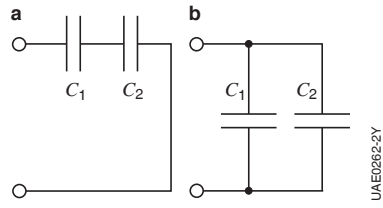
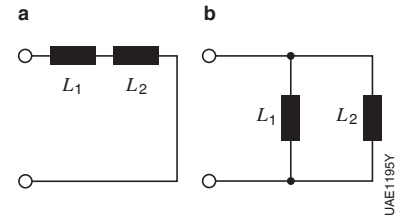


Figure 8: Connection of coils
a) Series connection,
b) Parallel connection.



Conversion of measures
Table 3 sets out the conversion factors which are to be applied when converting the different units.

Measurement of electric power
Example
The energy costs can be calculated from the electrical work and the applicable tariff of the energy supply company. The costs for the work and the costs at a tariff of 35 cents/kWh are to be calculated when U = 230 V, I = 5 A and t = 24 h.

Figure 9: Current and voltage measurement
R Load/consumer,
A Ammeter in current path,
V Shunt-connected voltmeter.

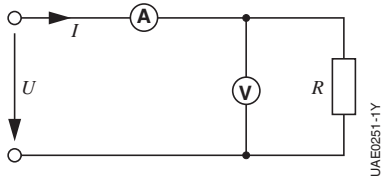
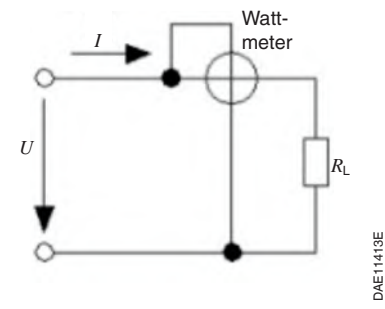


Figure 10: Connection of a wattmeter
U Voltage,
I Current,
R_L Load resistance.



$W = U\,I\,t$
= 230 V · 5 A · 24 h = 27,600 Wh
= 27.6 kWh.

Costs: 27.6 kWh · 35 cents/kWh = 9.66 €.

Example
In combination with a stopwatch the electricity meter also serves as a wattmeter:

$$P = \frac{W}{t}$$

In response to a heating appliance being switched on the meter reading changed from 1432.3 kWh to 1432.6 kWh in 18 minutes. The power of the heating appliance is obtained as follows:

P = 1432.6 kWh – 1432.3 kWh = 0.3 kWh
= 300 Wh = 18,000 Wmin.

$$P = \frac{W}{t} = 18{,}000\ \text{Wmin}/18\ \text{min} = 1000\ \text{W}.$$

The following base quantities are obtained from the International System of Units for direct conversions:

1 Ws = 1 V · 1 A · 1 s = 1 Nm = 1 kg · m^2/s^2.

Table 3: Conversion of units

	Ws	Wmin	Wh	kWh
Ws (watt-seconds)	1	1/60	$2.78 \cdot 10^{-4}$	$2.78 \cdot 10^{-7}$
Wmin (watt-minutes)	60	1	1/60	$1/60 \cdot 10^{-3}$
Wh (watt-hours)	$3.6 \cdot 10^3$	60	1	0.001
kWh (kilowatt-hours)	$3.6 \cdot 10^6$	$60 \cdot 10^3$	10^3	1

Time-dependent current

Charging and discharging a capacitor
A capacitor on direct voltage has two operating states: charging and discharging. When a (discharged) capacitor is connected to a direct-voltage source (Figure 11), it takes on charge, i.e., it is charged. The capacitor voltage rises over time from zero at the outset to approaching the applied direct voltage. A charging current flows which conversely assumes a peak value at the start and then decreases to zero.

The greater the resistance R through which the charging current flows and the greater the capacitance C (absorbable charge quantity), the longer a charging process takes. The product of these two quantities, the time constant τ, is the measure of the duration of the charging process.

$$\tau = R\,C.$$

Charging process

$$I = \frac{U_0}{R} e^{-\frac{t}{\tau}},$$

$$U = U_0 (1 - e^{-\frac{t}{\tau}}).$$

Discharging process

$$I = \frac{U_0}{R} e^{-\frac{t}{\tau}},$$

$$U = U_0\, e^{-\frac{t}{\tau}}.$$

U_0 Charging voltage or voltage at start of discharge,
I Charging or discharging current,
I_0 $I_0 = U_0/R$ Current at start of charge,
R Charging or discharging resistance,
U Capacitor voltage,
τ Time constant.

Charging and discharging currents have opposite directions.

Example 1
The capacitor can charge, C = 1 µF, via the resistance of R = 1 kΩ. Voltage U = 10 V. Charging time t = 2.5 ms. How high is the voltage U?

$\tau = 1\ \text{k}\Omega \cdot 1\ \mu\text{F} = 1$ ms.
$\tau/t =$ 1 ms/2.5 ms = 0.4.

$U = 10\ \text{V} \cdot (1 - e^{-0.4}) = 9.18$ V.

Example 2
The capacitor can charge with capacitance C = 0.15 µF via the resistance of R = 10 kΩ. Voltage U = 10 V. Charging time t = 1 ms. How high is the current I?

$\tau = 10\ \text{k}\Omega \cdot 0.15\ \mu\text{F} = 1.5$ ms.
$t/\tau =$ 1 ms/1.5 ms ≈ 0.667.

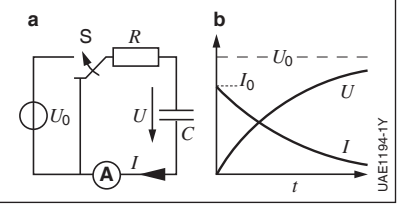
$$I = \frac{U_0}{R} e^{-0.667} = 10\ \text{V}/10\ \text{k}\Omega \cdot 0.513 = 0.513\ \text{mA}.$$

After the duration of a time constant of $t = 1\tau$ the capacitor voltage amounts to 63 % of the full charging voltage. After $t = 5\tau$ the capacitor voltage has risen to 99 % of the applied voltage, i.e., almost to the full value of U.

When a charged capacitor is connected to a resistance, it constitutes a voltage source in this circuit; the charges balance themselves out, a drawable discharging current flows, and the capacitor is discharged.

Capacitor in the alternating-current circuit
In the capacitor a current will only then flow when the effective voltage changes, e.g., when a direct voltage is switched on and off. When the capacitor voltage U_C has reached the value of the total voltage U, no further current flows. The capacitive re-

Figure 11: Charging process at capacitor
a) Circuit,
b) Voltage and current curves.
U Voltage,
I Current,
C Capacitance,
R Resistance,
A Ammeter.



sistance of the charged capacitor appears infinitely high.

In the alternating-current circuit a voltage whose direction changes periodically acts on the capacitor. Charging and discharging therefore alternate likewise in periodic sequence – this consideration should be confined to sinusoidal alternating voltage. The capacitor is charged during the positive half-wave and discharged again via the internal resistance of the voltage generator when the voltage has exceeded its peak value. The current direction is reversed in the process.

For the negative half-wave this process is repeated with the opposite preceding sign. The two capacitor plates alternately carry an electron surfeit and an electron deficit. A certain current intensity can be determined with a measuring device. The capacitor works as a resistor in the case of alternating voltage. The capacitor current (due to charging and discharging) then reaches its highest value when the greatest voltage change is present. Transferred to the sinusoidal alternating voltage, this means high current intensity at voltage zero. Current and voltage are in phase quadrature (Figure 12), the capacitor current leads the voltage at every instant by 1/4 period (1 period = 360°). Because current and voltage do not take effect at the same time, in contrast to ohmic resistance there is no thermal output in the capacitor. The capacitor's AC resistance is called capacitive reactance X_C.

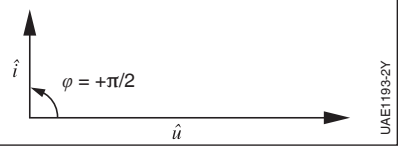
For an ideal capacitor the phase displacement angle φ between voltage and current is –90°. Because of the finite dielectric resistance, occurring surface currents and dielectric losses through polarity reversal in the case of alternating voltage, this angle is however not quite –90° for the real capacitor. This deviation is described by the loss factor or a reciprocal value, the capacitor Q factor. A capacitor separated from its voltage source will discharge slowly via its dielectric.

At a constant voltage the current intensity is all the greater the higher the frequency of the applied alternating voltage (with more charge reversals in a certain time) and the higher the capacitance C (because more electrons participate in the charge reversals). The current-determining capacitive reactance X_C is accordingly inversely proportional to the frequency f and the capacitance C. For sinusoidal alternating voltage the angular frequency $\omega = 2\pi f$ must be used.

$$X_C = 1/(2\pi f C) = 1/\omega C.$$

X_C Capacitive reactance in Ω,
f Frequency in Hz,
C Capacitance in F.

Figure 12: Phasor diagram, capacitor
$\hat{u}$ Voltage amplitude,
$\hat{i}$ Current amplitude.



Charging and discharging a coil

Switching on the coil

When the time characteristic of the making current of a coil is measured, a similar picture is obtained as for the time characteristic of the making voltage at a capacitance.

The cause of the delayed current rise is the inductance L of the coil. During switch-on the build-up of the coil's magnetic field starts. This field build-up is counteracted by the coil's self-induction voltage. Crucial to the charging process is the time constant $\tau = L/R$. Figure 13 shows the voltage and current characteristics.

$$U = U_0\, e^{-\frac{t}{\tau}}.$$

$$I = \frac{U_0}{R}\,(1 - e^{-\frac{t}{\tau}}).$$

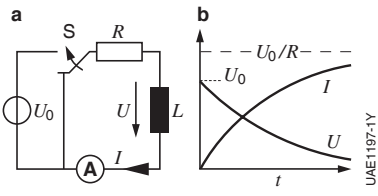
Switching off the coil
When an inductor is switched off in the direct-current circuit, the induction voltage is directed in such a way as to counteract the field decay. It is all the higher the faster the field decay (due to the change of current). When circuits with inductors are opened, high induction voltages are obtained due to the small time constant $\tau = L/R$ (R approaches infinity, τ thus approaches zero) because the rapidity of the current change during switch-off is particularly high. The induction voltage therefore drops from its peak value at the start of the breaking process in the time characteristic of an e function with the time constant $\tau = L/R$ to zero. The current during the breaking process also follows the same characteristic.

$$I = I_0\, e^{-\frac{t}{\tau}}.$$

$$U = I_0\, R\, e^{-\frac{t}{\tau}}.$$

U_0 Charging voltage or voltage at start of discharge,
I Instantaneous value of coil current,
I_0 $I_0 = U_0/R$ Current at making instant in A,
R Charging or discharging resistance,
U Capacitor voltage,
t Time after start of switch-on/off,
τ Time constant in s.

Figure 13: Making and breaking process of a coil
a) Circuit,
b) Voltage and current curves.
U Voltage,
U_0 Initial voltage,
I Current,
L Inductance,
R Resistance,
A Ammeter.



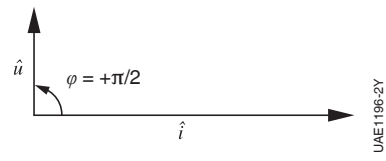
Coil in the alternating-current circuit
Inductive reactance
The coil current cannot follow rapid voltage changes because the inductance has a current-delaying effect. When the direct voltage is switched on, no current flows in the first moment. In the switch-off moment the voltage U reverts sharply to the value zero while the current intensity only decays with a delay. The coil current lags behind the voltage.

In the case of sinusoidal alternating voltage the inductance causes a time difference between identical phases of current and voltage. The current reaches for example the positive crest value in each case 1/4 period later than the voltage; the crest values of the coil voltage U_L include in each case the current intensity zero. The phase displacement between current and voltage is 90° if the coil losses (e.g., winding resistance) are disregarded (Figure 14). The curve shape of the two quantities remains unchanged here.

In coils the current delayed by the inductance cannot reach its full value because the crest values of the sinusoidal alternating voltage in contrast to direct voltage only occur as instantaneous values. The current intensity therefore complies with the applied alternating voltage and with the self-induction voltage generated in the coil. This influence of the inductance is summarized in the inductive reactance X_L. High frequencies result in rapid field changes in the coil and thus high self-induction voltages. The inductive reactance is therefore proportional to the inductance L and the frequency f.

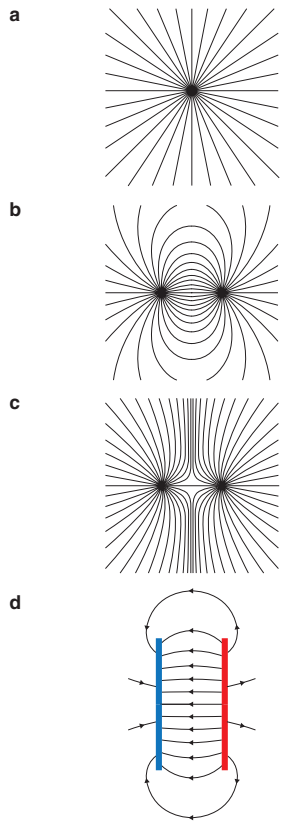
$$X_L = 2\pi f L = \omega L$$

Figure 14: Phasor diagram, coil
$\hat{u}$ Voltage amplitude,
$\hat{\imath}$ Current amplitude.



The factor 2π takes into account, as with the capacitive reactance, the time sinusoidal characteristic of current and voltage. Because current intensity and voltage are effective at different times and in different directions, no heat is obtained in the loss-free (ideal) coil. This fact explains the designation “reactance”.

Figure 15: Field pattern in an electric field (examples)
a) Individual point charge,
b) Two neighboring point charges with different charge,
c) Two neighboring point charges with identical charge,
d) Plate capacitor.



Electric field

Features

The force acting on a static electric charge is ascribed to the effect of an electric field. Figure 15 shows some field patterns by way of example. An electrostatic field can be defined by the following quantities.

Electric potential φ(P) and voltage U

The electric potential φ(P) at point P is a measure of the work required per charge to move the charge Q from a reference point to point P:

$$\varphi(\mathrm{P}) = \frac{W(\mathrm{P})}{Q} .$$

The voltage U is the potential difference (using the same reference point) between two points P_1 and P_2:

$$U = \varphi(\mathrm{P}_1) - \varphi(\mathrm{P}_2) .$$

Electric field strength E

The electric field strength E at point P depends on the location and its surrounding charges. It defines the maximum slope of the potential gradient at point P. The following applies to the field strength at a distance a from a positive point charge Q_1: It is directed away from the charge Q_1 and has the value

$$E = \frac{Q_1}{4\pi\varepsilon_0 a^2} .$$

Acting on a positive charge Q_2 at point P is a force in the direction of the electric field strength of the value

$$F = Q_2 E .$$

Electric field and matter

In a material which can be polarized (dielectric), an electric field generates electric dipoles (positive and negative charges $\pm Q$ at a distance a; Qa is called the dipole moment). The dipole moment per unit volume is called the polarization M. The displacement density D indicates the density of the electric displacement flux, The following applies:

$$D = \varepsilon E = \varepsilon_0 \varepsilon_\mathrm{r} E = \varepsilon_0 E + M .$$

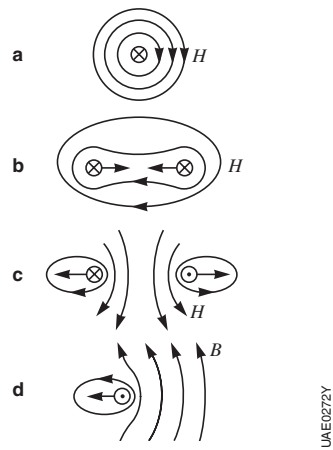
$\varepsilon = \varepsilon_0 \varepsilon_r$ is the dielectric constant of the material, ε_0 the electric field constant (dielectric constant of vacuum), ε_r the permittivity (relative dielectric constant). For air, $\varepsilon_r = 1$.

The quantity

$$w_e = \frac{1}{2} E D$$

is the electrical energy density. When multiplied by the volume, it produces the electrical energy W_e.

Figure 16: Current-carrying conductors and the associated magnetic lines of force
a) Single current-carrying conductor with magnetic field,
b) Parallel conductors, current flow in the same direction (conductors attract each other),
c) Parallel conductors, current flow in opposite directions (conductors repel each other),
d) A magnetic field of flux density B exerts a force on the current-carrying conductor (direction of force is determined using the right-hand rule).



Magnetic field

Features
Moving charges are associated with a magnetic field, i.e., electric current generates a magnetic field in the space surrounding it. Its manifestations match the manifestations of natural magnetic fields, e.g., the field of a permanent magnet or the Earth's magnetic field. In the same way as for electric fields, field lines were introduced to facilitate the depiction of magnetic fields (Figure 16). The direction in which the current flows (⊗ current flow into the page, ⊙ current flow out of the page) in the case of positive currents and the direction of the magnetic field strength form a right-handed screw. Whereas, however, in an electric field the electric field lines start out radially from the positively charged electrode and end at the negatively charged electrode, the magnetic field lines are always self-contained, i.e., they exhibit neither a start nor an end.

Another significant difference between an electric field and a magnetic field is that a magnetic field always has a pole character. Therefore it is not possible to realize a separate north or south pole. The field lines are always directed from the north to the south pole of the magnetic field and never touch.

Magnetic field quantities
A series of magnetic field quantities is required to calculate magnetic circuits.

Ampere-turns
In practice the magnetic fields of conductor loops and particularly of coils, which can be considered as the series connection of many conductor loops, are significant. These coils are as important as components in electrical engineering and electronics as resistors and capacitors. The connection between the current and the number of coil turns as the cause of the generated magnetic field is registered by the current linkage Θ (Figure 17).

$\Theta = NI$,

Θ Ampere-turns in A (magnetic potential),
N Number of turns on coil,
I Current in A.

Magnetic field strength
The magnetic field strength H assigns as a vector quantity to each point in space a strength and direction of the magnetic field generated by the magnetic potential (Figure 18).

$H = \Theta / l$,
$H = NI / l$

H Magnetic field strength in A/m,
l Mean field-line length in m,
N Number of turns.

Magnetic flux and flux density
The magnetic flux Φ is the measure of the sum total of magnetic field lines (Figure 19).

Figure 17: Magnetomotive force
I Current,
N Number of turns.

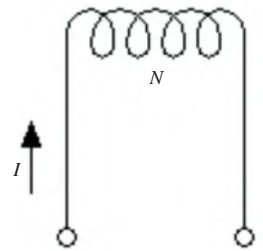
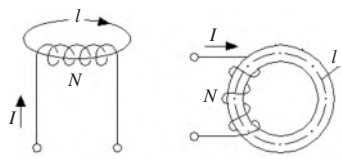


Figure 18: Magnetic field strength
I Current,
N Number of turns,
l Mean field-line length.



Another important field quantity is the flux density or magnetic induction B.

$B = \Phi / A$,
B Magnetic flux density in tesla
1 T = 1 tesla = 1 Vs/m^2,
Φ Magnetic flux in Weber (Vs)
1 Wb = 1 weber = 1 Vs,
A Area in m^2.

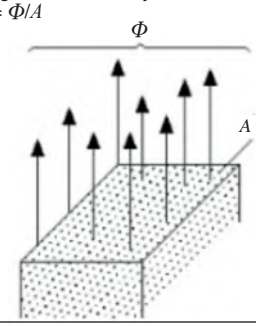
Magnetism and magnetic field
Magnets refer to bodies with the property of exerting a force on certain materials, e.g., iron, in its environment; a magnet attracts iron. This magnetic effect is intensified at the two ends of the magnet poles.

The ampere-turns Θ (current linkage or magnetomotive force) constitutes a measure of the magnetic effect of a current-carrying coil. Here the ampere-turns is all the greater the higher the current I and the number of turns N on the coil.

With complete magnetic circuits without considerable stray fields (ring coils, coils with a complete iron core, if necessary also with an air gap) the mean field-line length is obtained from the mean path length. It must be noted that for different materials or different cross-sections the following relationship applies:

$\Theta = I_1 N_1 + I_2 N_2 + \ldots$

Figure 19: Magnetic flux and magnetic flux density (induction)
Φ Magnetic flux,
A Area,
B Magnetic flux density.
$B = \Phi / A$



The magnetic field of a current-carrying coil can be amplified by a multiple when a ferromagnetic material – e.g., in the form of an iron core – is introduced into the coil cavity. This influence is taken into account in the magnetic flux density B, whereby the causative field strength H is multiplied by a factor, the so-called permeability μ.

$$B = \mu H.$$

Here the permeability μ is made up of two components:
- the magnetic field constant $\mu_0 \approx 1.256 \cdot 10^{-6}$ Vs/Am (the value for vacuum), and
- the relative permeability μ_r, which as a pure numerical factor specifies by how much the magnetic flux density is increased by a specific material in the coil cavity as against vacuum.

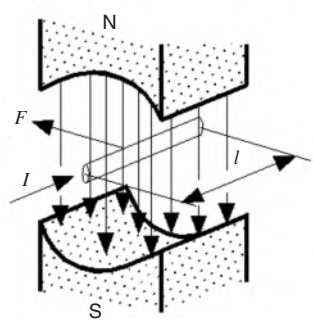
With $\mu = \mu_0 \mu_r$, therefore, the magnetic flux density is obtained

$$B = \mu_0 \mu_r H \approx 1.256 \cdot 10^{-6} \text{ Vs/Am } \mu_r H.$$

With the magnetic field strength H in A/m and μ_0 in Vs/Am the unit Vs/m^2 is obtained for the magnetic flux density B and is denoted by tesla (T).

With B in Vs/m^2 and A in m^2 the unit Vs is obtained for Φ and is abbreviated to weber (Wb).

Figure 20: Current-carrying conductor in a magnetic field
I Current,
F Force on the conductor,
l Conductor length in magnetic field.



Forces on current-carrying conductors in the magnetic field

Magnetic fields exert forces on each other. Unlike poles attract, like poles repel. These effects can be substantiated with permanent magnets (e.g., magnetic bar and magnetic needle). Even the current-carrying conductor with its magnetic field is exposed to forces by other magnetic fields (Figure 20). The current-carrying conductor amplifies the right part of the permanent-magnet field (same field-line direction). To the left of the conductor the permanent-magnet field is weakened (opposite field-line direction).

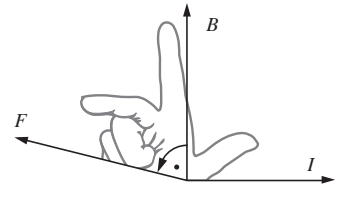
The current-carrying conductor thus experiences a force in the permanent magnet's magnetic field. The force F is obtained as

$$F = B I l z \sin\alpha.$$

F Force on the conductor in N (newton),
B Magnetic flux density in T (tesla),
l Effective conductor length in m (meter),
α Angle between l and B, in Figure 20 $\sin\alpha = \sin 90° = 1$.
z Number of conductors (number of turns),
I Current intensity in A (ampere).

The direction of the force can be determined by applying the right-hand rule (Figure 21). The thumb points in the direction of the positive current flow (cause), the index finger points in the direction of the magnetic field (relaying), and finally the middle finger specifies the direction of force (effect).

Figure 21: Right-hand rule
I Current,
B Magnetic flux density,
F Force.



Electrodynamic principle

Induction

Whereas in a motor electrical energy is converted into mechanical energy, with a generator mechanical energy can be converted into electrical energy. If namely a conductor is moved in a magnetic field in such a way that it intersects field lines, a voltage is induced in it while it is moving (Figure 22). The level of the induced voltage U_{ind} depends here on the magnitude of the magnetic flux density B, the rate of motion v of the conductor loops, and thus on the change of area dA in the time unit dt as well as on the number of turns z. The following applies:

$$U_{ind} = -z\,B\,dA/dt.$$

Self-induction

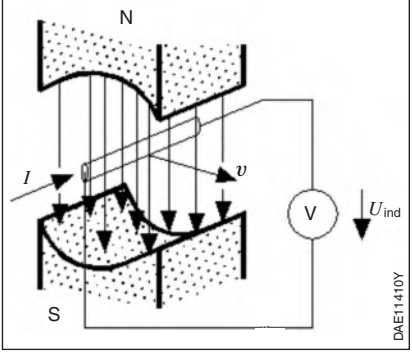
When the current intensity and thus the magnetic field change in a conductor or in a coil, an induction voltage is also generated in the conductor or coil itself. The magnitude of this self-induction voltage depends just like the magnitude of the induction voltage (of the rest or the motion) on the extent and rapidity of the magnet-field change.

The property of a coil of generating a certain self-induction voltage U_{ind} for a certain current change dI per time unit dt is known as inductance L. Its unit is obtained from the generated voltage (in volts) per current change (in A/s) as V/(A/s) = Vs/A, abbreviated as henry (H). The following applies:

$$U_{ind} = -L\,(dI/dt).$$

In all cases the level of the generated induction voltage depends on the extent to which the strength of the magnetic field (B or Φ) and the rapidity with which the strength of the magnetic field changes.

Figure 22: Induction of motion
I Current,
v Speed of the conductor in the magnetic field perpendicular to the field lines,
U_{ind} Induced voltage.



Magnetic field and matter

Magnetic polarization

In matter, induction B theoretically consists of two components. One component comes from the applied field ($\mu_0 H$), and the other from the matter (J):

$$B = \mu_0 H + J.$$

J is the magnetic polarization and describes that component of flux density contributed by the matter. In physical terms, J corresponds to one magnetic dipole moment per unit volume and is generally a function of field strength H. For many materials $J >> \mu_0 H$ and is proportional to H. Thus:

$$B = \mu_r \mu_0 H$$

with relative permeability μ_r; in free space it has the value $\mu_r = 1$.

Magnetic energy density

The quantity

$$w_m = \frac{1}{2} B H$$

is called magnetic energy density. When multiplied by the volume, it produces the magnetic energy W_m.

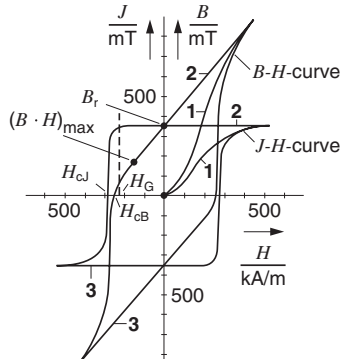
Classification of materials
Materials are divided into three groups according to their relative permeability values.

Diamagnetic materials
μ_r is independent of magnetic field strength and smaller than 1; the values are within the range:
$(1-10^{-5}) < \mu_r < (1-10^{-11})$
(e.g., Ag, Au, Cd, Cu, Hg, Pb, Zn, water, organic materials, gases).

Paramagnetic materials
μ_r is independent of magnetic field strength and greater than 1; the values are within the range
$(1+10^{-8}) < \mu_r < (1+10^{-4})$
(e.g., O_2, Al, Pt, Ti).

Ferromagnetic materials
The magnetic polarization in these materials is very high, and its change as a function of the field strength H is nonlinear; it is also dependent on hysteresis. Nevertheless, if, as is usual in electrical engineering, the relationship $B=\mu_r\mu_0 H$ is chosen, then μ_r is a function of H and exhibits hysteresis; the values for μ_r are within the range
$10^2 < \mu_r < 5\cdot 10^5$
(e.g., Fe, Co, Ni, ferrites).

Figure 23: Hysteresis loop (e.g., hard ferrite)
1 Rise path,
2, 3 Demagnetization curves.
H Magnetic field strength,
B Magnetic flux density,
J Magnetic polarization,
B_r Remanence,
H_{cB}, H_{cJ} Coercive field strength,
H_G Limiting field strength.



Hysteresis loop
The hysteresis loop (Figure 23), which illustrates the relationship between B and H as well as J and H, is explained as follows: If the material is in the unmagnetized state ($B=J=0$, $H=0$) when a magnetic field H is applied, the magnetization follows the rise path (1). From a specific, material-dependent field strength, all magnetic dipoles are aligned and J reaches the value of saturation polarization J_s (material-dependent) which can no longer be increased. If H is now reduced, J decreases along section (2) of the curve and at $H=0$ intersects the B or J axis at the remanence point B_r or J_r (in which case $B_r=J_r$). The flux density and polarization drop to zero only on application of an opposing field whose field strength is H_{cB} or H_{cJ}; this field strength is called the coercive field strength.

As the field strength of the opposing field is further increased, saturation polarization is reached in the opposite direction. If the field strength is again reduced and the field reversed, curve (3), which is symmetrical to curve section (2), is traversed.

The most important characteristic values of a ferromagnetic material are usually tabulated:
- saturation polarization J_s,
- remanence B_r (residual induction for $H=0$),
- coercive field strength H_{cB} (demagnetizing field strength where B becomes equal to 0) and
- coercive field strength H_{cJ} (demagnetizing field strength where J becomes equal to 0, of significance only for permanent magnets),
- limiting field strength H_G (a permanent magnet remains stable up to this field strength),
- maximum small-signal permeability μ_{max} (maximum slope of the rise path; significant only for soft magnetic materials),
- hysteresis loss (energy converted into heat per volume during one remagnetizing cycle, corresponds to the area of the B–H hysteresis loop; significant only for soft magnetic materials).

Ferromagnetic materials
Ferromagnetic materials are divided into soft and permanent magnetic materials. What must be emphasized is the immense range of eight powers of ten covered by the coercive field strength.

Permanent-magnet materials
Permanent-magnet materials have high coercive field strengths; the values lie within the range

$$H_{cJ} > 1 \frac{kA}{m}$$

Thus high demagnetizing fields H may occur without the material losing its magnetic polarization. The magnetic state and operating range of a permanent magnet lie within the second quadrant of the hysteresis loop, on the demagnetization curve.

In practice, the operating point of a permanent magnet never coincides with the remanence point, because magnetic polarization of the permanent magnet in its interior always causes a magnetic field, the demagnetizing field, which shifts the operating point into the second quadrant.

The point on the demagnetization curve at which the product BH reaches its maximum value $(BH)_{max}$ is a measure of the maximum attainable air-gap energy. In addition to remanence and coercive field strength, this value is important for characterizing permanent magnets.

AlNiCo, ferrite, FeNdB (REFe), and SeCo magnets are currently the most important types of permanent magnets in terms of industrial applications; their demagnetization curves (Figure 24) exhibit characteristics typical of the individual magnet types.

Soft magnetic materials
Soft magnetic materials have a low coercive field strength

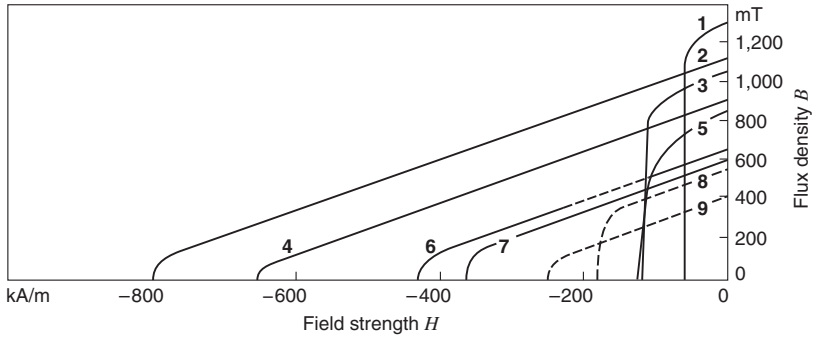
$$H_{cJ} < 1 \frac{kA}{m},$$

i.e., a narrow hysteresis loop. The flux density assumes high values (large μ_r values) for low field strengths so that, in customary applications $J >> \mu_0 H$, i.e., in practice no distinction need be made between $B(H)$- and $J(H)$ curves.

Due to their high induction at low field strengths, soft magnetic materials are used as conductors of magnetic flux. As they exhibit low remagnetization losses (hysteresis loss), materials with low coercive field strengths are particularly well-suited for applications in alternating magnetic fields.

The characteristics of soft magnetic materials depend essentially on their pretreatment. Machining increases the coercive field strength, i.e., the hysteresis loop becomes broader. The coercive field strength can be subsequently reduced to its initial value through material-specific annealing at high temperatures (magnetic

Figure 24: Demagnetization curves for various permanent-magnet materials
1 AlNiCo 52/6, 2 REFe 220/140, 3 AlNiCo 60/11, 4 SECo 112/100, 5 AlNiCo 30/10, 6 SECo 70/70p, 7 PlCo 60/40, 8 MnAl, 9 Hard ferrite 25/25.



final annealing). The magnetization curves, i.e., the B-H relationships, are set out in Figure 25 for several important soft magnetic materials.

Remagnetization losses
In Table 4 P1 and P1.5 represent the remagnetization loss for inductions of 1 and 1.5 tesla respectively, in a 50 Hz field at 20 °C. These losses are composed of hysteresis losses and eddy-current losses. The eddy-current losses are caused by electric fields which are induced (law of induction) in the magnetically soft components of the magnetic circuit as a result of changes in flux during alternating-field magnetization. Eddy-current losses can be kept low by applying the following measures to reduce electric conductivity:

– lamination of the core,
– use of alloyed materials (e.g., silicon iron),
– use of insulated powder particles (powdered cores) in the higher frequency range,
– use of ceramic materials (ferrites).

Figure 25: Magnetization curves for soft magnetic materials
1 Pure iron, 2 78 NiFe (Permalloy), 3 36 NiFe, 4 Ni-Zn ferrite, 5 50 CoFe, 6 V360-50A (electrical sheet steel), 7 Structural steel, 8 Cast iron, 9 Powder Fe core.

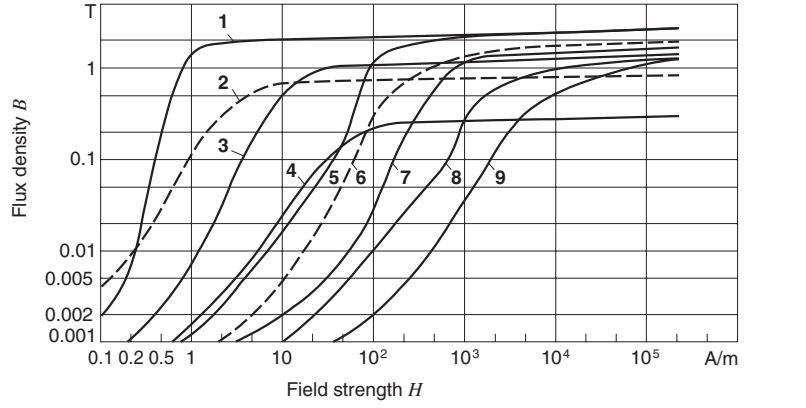


Table 4: Remagnetization losses

Sheet grade	Nominal thickness mm	Remagnetization loss W/kg P1	P 1.5	B (for H = 10 kA/m) T
M 270 – 35 A	0.35	1.1	2.7	1.70
M 330 – 35 A	0.35	1.3	33.3	1.70
M 400 – 50 A	0.5	1.7	4.0	1.71
M 530 – 50 A	0.5	2.3	5.3	1.74
M 800 – 50 A	0.5	3.6	8.1	1.77

Electromagnetic fields

Electrical engineering deals with electromagnetic fields and their effects. These fields are associated with electric charges (in each case an integral multiple of the electric elementary charge). Physics does not indicate whether the fields are the cause or the effect. Static charges produce an electric field, whereas moving charges give rise to a magnetic field as well. The association of electric and magnetic fields with static and moving charges is described by Maxwell's equations [3].

The presence of these fields is evidenced by the effects of their forces on other electric charges. The force on a point charge Q in an electric field is called Coulomb force. It causes a repulsion between charges of the same name. For two point charges Q_1 and Q_2 in free space at a distance a, it is:

$$F = \frac{Q_1 Q_2}{4\pi\varepsilon_0 a^2} .$$

$\varepsilon_0 \approx 8.854 \cdot 10^{-12}$ F/m is the electric field constant, also called the dielectric constant of free space.

The force acting on a moving charge in a magnetic field is expressed by the Lorentz force. This is responsible for two parallel conductors which carry rectified currents I_1 and I_2 being mutually attracting. Over a length l in free space at conductor distance a the force of attraction between the two conductors is:

$$F = \frac{\mu_0 I_1 I_2 l}{2\pi a} .$$

$\mu_0 \approx 1.257 \cdot 10^{-6}$ H/m stands for the magnetic field constant, also called the permittivity of free space.

Power in the alternating-current circuit

Features

Resistors, capacitors, and coils are important components in electrical engineering and electronics. Especially in electronic circuits they can be found in the most varied combinations and circuit variants. But also many electrical devices exhibit as consumers (e.g., motors and fluorescent lamps) behavior that can be emulated with an equivalent circuit of resistors, capacitors, and coils.

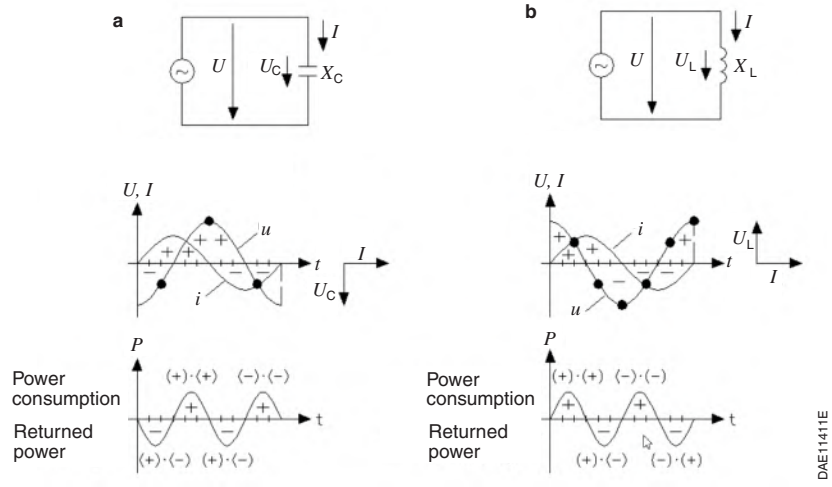
The behavior of these components and their interaction are of significance particularly in the case of operation on sinusoidal alternating voltage. If an ohmic resistance is connected to a sinusoidal voltage, current and voltage are constantly in phase. For a capacitor the current leads the voltage by 90°, while for a coil the current lags behind the voltage by 90°.

The occurring phase displacements can be depicted in both line and phasor diagrams. However, the phase diagrams for voltages, currents, resistances, and power are particularly important to describing the interaction of active resistances and reactances. In this connection a right-angled triangle appears in each phasor diagram so that mathematical relationships can be derived using Pythagoras' theorem. But the trigonometric functions sine, cosine, and tangent can also be used to describe the mathematical relationships in phasor diagrams and thus in alternating-current circuits.

There are three different types of power in an alternating-current circuit: Active power P in watts, apparent power S in va (volt-ampere), and reactive power Q in var (volt-ampere-reactive).

Figure 26: Comparison of capacitive and inductive reactive power

Capacitor and coil are taken as ideal, resulting in phase quadrature of current and voltage.
a) Current, voltage, and power diagram for capacitive reactive power,
b) Current, voltage, and power diagram for inductive reactive power.



Active power
The ohmic resistance in a heater consumes in alternating current pure active power because current I and voltage U are in phase. Multiplying the instantaneous values of voltage and current produces the instantaneous values of power $P = U\,I$.

Example:
An electric heating plate is connected to U = 230 V and a current of I = 3 A flows. How high is the active power?

$P = UI$ = 230 V · 3 A = 690 W

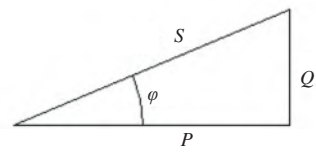
Reactive power
Owing to the phase displacement between voltage and current u and i do not always have the same preceding sign (Figure 26). For the power an area is obtained which is situated precisely halfway each in the positive and negative regions. "Positive" power is the power drawn by the consumer from the voltage source.

The values U and I are multiplied and produce the reactive power $Q = U\,I \sin\alpha$ or $Q = S \sin\alpha$.

Apparent power
An apparent power S is created when effective resistances (R) and reactances (X_L or X_C) are interconnected. Multiplying for such a circuit the measured voltage U by the measured current I produces the apparent power S.

The apparent power does not permit any conclusion as to how big the share of the actually drawn and "consumed" power is.

Figure 27: Active, apparent, and reactive power in power triangle
P Active power,
Q Reactive power,
S Apparent power.



Active and reactive power can be added to the apparent power in the phasor diagram (Figure 27).

For a sinusoidal characteristic $S = UI$. The phase displacement φ of current and voltage can be calculated

$P = UI\cos\varphi$
$P = S\cos\varphi$
$Q = UI\sin\varphi$

S Apparent power in VA,
P Active power in W,
Q Reactive power in var,
U Voltage (effective value),
I Current (effective value),
cos φ Power factor,
sin φ Reactive-power factor.

Reactive-power compensation
The energy supplier companies make it compulsory for their customers to achieve a power factor of around cos φ = 0.9 in their systems. If this value is undershot, the phase angle is reduced by individual or group compensation. By connecting an appropriate capacitor in parallel, it is possible to free the incoming cable in part or in whole of reactive current. In this way, the reactive currents or reactive power in the incoming cable cancel each other out and these can be output with active current or active power.

Wave propagation

Waveguide

At higher frequencies forward and return conductors of an electric connection section are combined and called a waveguide or in short a line. Common structures are the parallel-wire line, the twisted two-wire line (twisted pair), the coaxial line, the strip line, and the microstrip line.

At higher frequencies the current is no longer distributed evenly over the line cross-section. In the individual wire the current is increasingly displaced to the edge as the frequency increases (skin effect). In lines there is an additional displacement of current to the other conductor in each case (proximity effect). In particular the ohmic loss resistance and thus the damping of the line are thereby increased.

A line has the characteristic impedance Z if it is terminated with this resistance Z and then demonstrates the same impedance Z as the input resistance. In the lines used in motor vehicles (TEM or L lines) the characteristic impedance Z is directly connected with the capacitance C per length l and the inductance L per length l:

$$Z = \sqrt{\frac{L}{C}} .$$

Table 5 shows characteristic impedances for some selected line types.

Even the amplitudes of the electric and magnetic fields along the line are linked via the characteristic impedance:

$$Z = \frac{\hat{E}}{\hat{H}} .$$

A line of characteristic impedance Z which is terminated with resistance R demonstrates at its end the reflectance factor

$$r = \frac{R - Z}{R + Z} .$$

The square of the absolute value of the reflectance factor indicates what fraction of the power is sent back at the end of the line.

$$P_r = |r|^2 P_0 .$$

Passed on to the terminating resistor, however, the power becomes

$$P_t = (1 - |r|^2) P_0 .$$

When $R = Z$ applies, adaptation occurs and the reflectance factor r is zero. In this case, no power is reflected, the full power being passed on to the terminating resistor R.

At higher frequencies and with longer lines, voltage and current along the line are no longer constant. At a fixed instant both vary along the line. Here there can be points over the length of the line at which the amplitude of the voltage fluctuations is at maximum, while at other points only minimum voltage amplitudes occur. The ratio of maximum to minimum amplitude on a line is called the standing wave ratio s.

$$s = \frac{U_{max}}{U_{min}} .$$

If a line is supplied at one end only with a sinusoidal time function, while terminated at the other end with a resistor, it is possible to calculate the standing wave ratio from the reflectance factor.

$$s = \frac{1 + |r|}{1 - |r|} .$$

Table 5: Characteristic impedances for some selected line types

Parallel conductors (twin conductors) in air	$Z = \sqrt{\frac{\mu_0}{\varepsilon_0}} \frac{1}{\pi} \ln \frac{a + \sqrt{a^2 - 4r^2}}{2r}$	a r	Distance between conductors [m] Conductor radius [m]
Concentric line (coaxial line)	$Z = \sqrt{\frac{\mu_r \mu_0}{\varepsilon_r \varepsilon_0}} \frac{1}{2\pi} \ln \frac{r_2}{r_1}$	r_2 r_1	Inside radius of outer tube [m] Radius of center tube [m]
Conductor to ground in air	$Z = \sqrt{\frac{\mu_0}{\varepsilon_0}} \frac{1}{2\pi} \ln \frac{a + \sqrt{a^2 - r^2}}{r}$	a r	Distance from conductor to ground [m] Conductor radius [m]

During adaptation s = 1 and at all points on the line the voltage amplitude is the same.

Couplers and matching transformers
Couplers are used in conjunction with matching transformers to forward an electromagnetic wave from one waveguide to another. Matching transformers reduce the reflectance factor during the transition from one characteristic impedance to another or couple a symmetrical waveguide (e.g. a parallel-wire line) with an asymmetrical one (e.g. a coaxial line).

Couplers and matching transformers are usually reciprocal, coupling just as well in one direction as in the other.

Wave propagation in free space
Even free space is a waveguide, with the characteristic impedance of free space (field characteristic impedance Z_0):

$$Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} \approx 377\ \Omega.$$

Far away from antenna structures the direction of the electric field, the direction of the magnetic field and the direction of propagation of the electromagnetic wave in free space are vertically on top of one another, and the electric field strength is linked with the magnetic field strength via the characteristic impedance Z_0. The electromagnetic power density S emitted by a transmitter is established via the amplitudes of the electric field strength $\hat{E}$ and the magnetic field strength $\hat{H}$. In free space, far away from the transmitter, the electromagnetic power density decreases with the inverse square of the distance from the transmitter:

$$S = \frac{1}{2}\hat{E}\,\hat{H},$$

$$S(r) = S(r_0)\,\frac{r_0^2}{r^2}.$$

Electric and magnetic field strengths, on the other hand, decrease only inversely with the distance from the transmitter:

$$\hat{E}(r) = \hat{E}(r_0)\,\frac{r_0}{r},$$

$$\hat{H}(r) = \hat{H}(r_0)\,\frac{r_0}{r}.$$

For some applications in motor vehicles the receiver is situated so near to the transmitter that near-field conditions prevail. Radio remote controls, for example, generate via a current-carrying coil magnetic fields whose field strength in the near field is inversely proportional to the cube of the distance from the transmitter.

Antennas
A coupler from a waveguide to free space is called an antenna. Antennas too are usually reciprocal; they can therefore act equally as transmitting and receiving antennas.

An electromagnetic wave in free space consists of time-variant electric and magnetic fields. These have one spatial direction – far away from transmitting antennas the field directions of these two fields are vertical to the direction of propagation of the wave. The direction of the corresponding electric field determines the polarization direction of the electromagnetic wave. Most antennas are designed in such a way that they can transmit and receive only one single direction of the electric (sometimes also the magnetic) field strength. They therefore transmit only linearly polarized waves and receive only one particular polarization of the wave.

A distinction is made between narrow-band antennas (dipole antennas, as used for example for GPS, GSM, Bluetooth, WLAN or even VHF) and wide-band antennas, which are used as transmitting and receiving antennas in EMC measurements in motor vehicles. An example of a narrow-band antenna is the half-wave dipole, a rod or wire of length l = 0.96 $\lambda/2$ (λ: wavelength in free space). The optimum length becomes shorter for thicker rods, or if necessary also through connection with additional capacitances. In conjunction with an extended metal surface the dipole length is shortened by half to one $\lambda/4$-dipole.

An example of a wide-band antenna is the logarithmically periodic antenna. Here, several half-wave dipoles are of different lengths are arranged along a twin line. No antenna transmits in transmitting mode uniformly in all directions. Likewise, no antenna receives in receiving mode all directions equally well. This directionality is described by the directional diagram (also known as the antenna directional characteristic). Figure 28 shows the directional diagram of a half-wave dipole in a vertical section through the dipole antenna. What is depicted is the radiated power density against the angle.

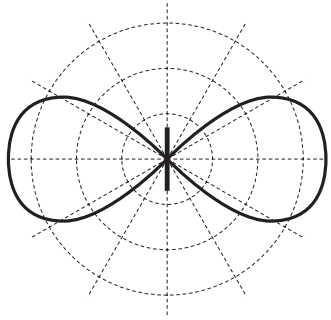
Often what is of interest is not the entire directional diagram, but only its maximum value in the preferred direction. The directivity indicates by how much the radiated power per surface area S_{max} of the antenna in transmitting mode is greater in the preferred direction than for a reference antenna which does not prefer a direction:

$$D_0 = \frac{S_{max}}{S_0}.$$

The directivity in relation to the antenna gain G is given as a dB value:

$$G = 10 \log D_0.$$

Figure 28: Directional diagram of a half-wave dipole
(Vertical section through the dipole antenna, power density against the angle).



The antenna losses are usually also included in the antenna gain such that G is smaller than in the ideal case. The same directivity and the same antenna gain also apply in receiving mode. There they indicate by how much the antenna power input is greater than for the reference antenna.

In the EMC standards calculations are made not with the antenna gain but with the antenna factor AF:

$$AF = 20 \log \frac{E}{U}.$$

The antenna factor indicates what voltage amplitude U (in V) drops at the input resistance of the measuring device connected to the antenna when an electric field of amplitude E (in V/m) is received. The same antenna factor also applies to transmitting mode and links the voltage amplitude U (in V) at the input of the transmitting antenna with the amplitude E (in V/m) of the electric field of the emitted wave.

Antenna gain and antenna factor can be converted into one another:

$$AF = 10 \log \left(\frac{4\pi Z_0}{\lambda^2 R_L} \right) - G$$

where
R_L Input resistance of measuring device and of antenna input (in Ω, usually 50 Ω), Z_0 Field characteristic impedance of free space (equal to 377 Ω), λ Wavelength (in m).

Electric effects in metallic conductors

Contact potential between conductors
Contact potential occurs in conductors, and is analogous to the triboelectricity or voltaic emf in insulators (e.g., glass, hard rubber). If two dissimilar metals at the same temperature are joined to make metal-to-metal contact with one another and are then separated, a contact potential is present between them. This is caused by the different work functions of the electrons. The magnitude of contact potential depends on the element positions in the thermoelectric series (Table 6). If more than two conductors are so joined, the resulting contact potential is the sum of the individual contact potential values.

Thermoelectricity and thermocouples
Seebeck effect
A voltage occurs between the two ends of a conductor if these exhibit different temperatures. As the temperature increases the electron density in the conductor decreases on account of the increasing kinetic energy of the electrons: The electrons in the "hot" zone have a greater kinetic energy and move on average towards the "cold" zone. In all, there is then a higher electron density in the "cold" zone than in the "hot" zone and hence a negative potential. In semiconductor physics this can be expressed by different Fermi levels and a resulting potential difference (thermoelectromotive force U_T). In all, the thermoelectromotive force U_T between the ends is obtained with the temperature difference ΔT according to the following equation:

$$U_T = \alpha \Delta T,$$

where
α: Seebeck coefficient in µV/K,
Δt: Temperature difference in K.

The Seebeck coefficient α is approximately temperature-independent and is determined by the material. The Seebeck coefficient α can be used to determine the thermoelectromotive force U_T in µV for the respective material for a given temperature difference in kelvin; frequently the second conductor is made of platinum – therefore the Seebeck coefficient α for platinum in Table 7 equals zero. Occasionally the k factor is used instead of α.

Technically, this effect cannot be used with a homogeneous material. Since the two measuring points are brought to the same temperature for the voltage tap via the measuring device, two equal voltages would be connected in series in the opposite direction. Consequently, no voltage would be tapped between the two mea-

Table 6: Placement of some elements in the thermoelectric series at the temperature 0 °C (source [4])
(For Pb the thermoelectromotive force is arbitrarily set at zero)

Element	Thermoelectromotive force [µV K^{-1}]
Antimony	+35
Iron	+16
Zinc	+3
Copper	+2.8
Silver	+2.7
Lead	0
Aluminum	–0.5
Platinum	–3.1
Nickel	–19
Bismuth	–70

Table 7: Seebeck coefficient of some elements referred to platinum at the standard temperature 0 °C (source [5])

Element	Thermoelectromotive force [µV K^{-1}]
Selenium	900
Tellurium	500
Silicon	440
Germanium	300
Antimony	47
Nickel chromium	25
Iron	19
Tungsten	7.5
Cadmium	7.5
Silver	6.5
Gold	6.5
Copper	6.5
Rhodium	6
Tantalum	4.5
Lead	4
Aluminum	3.5
Carbon	3
Mercury	0.6
Platinum	0
Sodium	–2
Potassium	–9
Nickel	–15
Constantan	–35
Bismuth	–72

suring points. This changes when different materials are used (Figure 29). The actual thermocouple consists in Figure 21 of two conductors made of materials 1 and 2. Since the connections are in some cases too short, they are extended at a coupling point. This coupling point at the temperature T_K. If necessary, a further extension is required – this is depicted in Figure 29 with material 3. This can involve, for example, conductor tracks of copper. Finally, the thermoelectromotive force U_T can be tapped. The electronic circuit here is at the ambient temperature T_U.

The series connection of the different elements in Figure 29 produces the following relationship:

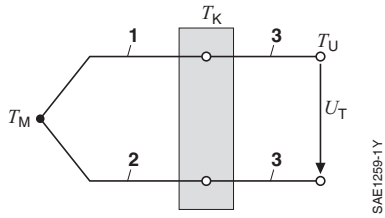
$$U_T = \alpha_3 (T_U - T_K) + \alpha_1 (T_K - T_M) + \alpha_2 (T_M - T_K) + \alpha_3 (T_K - T_U),$$

$$U_T = (\alpha_1 - \alpha_2) (T_K - T_M).$$

To match the material pairing of a thermocouple corresponding extensions of the same material can be used as compensating lines.

The advantage of thermocouples is their small size, enabling them to be used in areas with very small dimensions – for example, thermocouples can be stuck to the surface for flow measurements in a wind tunnel. The thermoelectromotive forces are indeed very low; however, the internal resistance of the thermocouples is low such that disturbances have little bearing on the output signal. Thermocouples can be used to good effect up to temperatures in excess of 2,000 °C. On the down side, however, thermocouples are subject to low long-term stability.

Figure 29: Basic design of a thermocouple
1 Conductor with material 1,
2 Conductor with material 2,
3 Conductor extension with material 3,
T_M Temperature at the measuring point,
T_K Temperature at the coupling point,
T_U Ambient temperature,
U_T Thermoelectromotive force.



Peltier effect
The reciprocal of the Seebeck effect is the Peltier effect, in which a temperature difference is created by electrical energy (e.g., as a heat pump). A current flow through the contact points of two different conductors results in heat energy Q being absorbed at one contact and picked off at the other. Depending on the current direction, heat can thus be supplied or dissipated at a contact. The heat transfer ΔQ per time unit Δt is governed by the following equation:

$$\frac{\Delta Q}{\Delta t} = \Pi I \text{ (in V A)},$$

where
Π: Peltier coefficient (in V),
I: Current intensity (in A),
Δt: Time interval (in s).

The relationship between the Peltier coefficient Π, the temperature T and the thermoelectromotive force U_T is as follows:

$$\Pi = U_T T .$$

To utilize the Peltier effect, bismuth telluride (Bi_2Te_3) or even silicon-germanium (SiGe) for example is used; these are either n- or p-doped so as to obtain zones with different Fermi levels. The maximum voltage per thermocouple (i.e., per n- and p-doped zone) is 0.12 V so that a Peltier element with 127 thermocouples requires a supply voltage of approximately 15 V.

It must be noted when operating a Peltier element that the cooling capacity does indeed increase with increasing current I, but this results on account of the electric (ohmic) power loss $P = RI^2$ in self-heating, which counteracts the cooling capacity. Consequently, determining the operating point with maximum efficiency is governed by the following: The ratio between current intensity and maximum current intensity must be chosen in the same way as that between the temperature difference and the maximum achievable temperature difference.

Galvanomagnetic and thermomagnetic effects

Such effects are understood to be changes caused by a magnetic field in the flow of electricity or heat within a conductor. There are twelve different recognized effects which fall into this category, the most well-known of which are the Hall, Ettingshausen, Righi-Leduc, and Nernst effects.

Hall effect

The Hall effect is of particular significance in technical applications. If a current I is sent through a suitable conductor and simultaneously a magnetic field or a magnetic flux density B is applied perpendicular to this current, a voltage is produced which is perpendicular to both the flow of current and the magnetic field. This voltage is called the Hall voltage U_H (Figure 30):

$$U_H = \frac{R_H I B}{d},$$

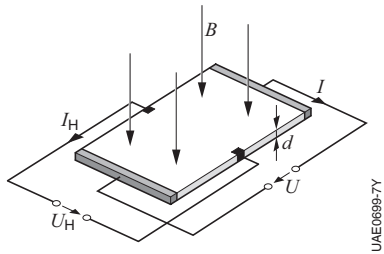
where

R_H Hall constant (in m^3/As),
I Supply current (in A),
B Magnetic flux density (in Vs/m^2),
d Thickness of conductor (in m).

In the case of Hall-effect sensors made of doped materials such as silicon – i.e., n- or p-type silicon – the majorities determine the sign and the direction of the Hall voltage U_H.

Figure 30: Hall-effect sensor element

B Flux density,
I Supply current,
U Supply voltage,
I_H Hall current,
U_H Hall voltage,
d Thickness of conductor.



The Hall constant R_H is relevant to use: Highest values in the range of 10^{-4} m^3/As are achieved for indium arsenide (InAs); for silicon, the values are among other things dependent on the doping and are much lower. When the current I is known, it is possible to infer the magnetic flux density B or the magnetic field. The advantage here is the strong linear dependence of the Hall voltage on the magnetic field or the magnetic flux density; the disadvantage is as a rule the not insignificant offset on account of inhomogeneities (crystal defects, particularly on the surface) and geometry influences.

In ferromagnetic materials, the Hall voltage is a function of magnetization, where hysteresis effects can also occur.

Hall-effect sensors are frequently used for rotational-speed recording, in which the alternation of tooth and gap of a rotating soft-magnetic sensor wheel is detected. Other applications in automobiles include the detection of end points or the potential-free (i.e., contactless) measurement of electric currents via the magnetic field generated by the current.

Magnetoresistive effects

Anisotropic magnetoresistive effect

Sensors based on the anisotropic magnetoresistive effect (AMR effect) change the electrical resistance R on account of an applied magnetic field H; in contrast to conventional Hall-effect sensors and also magnetoresistors, AMR sensors react to magnetic fields in the layer plane (here xy plane in Figure 31).

AMR sensor elements essentially consist of a thin strip roughly only 20 to 50 nm thick of highly permeable permalloy ($Ni_{81}Fe_{19}$ alloy), which exhibits strong shape anisotropy (Figure 31). This is expressed by a much greater length l than width b; this defines an "easy axis", which determines the orientation of magnetization M without an external magnetic field H in the xy plane in Figure 31. A magnetic field H in the direction of the "hard axis", in Figure 31 in the y direction) rotates M out of this position, which results in a relative change of the electrical resistance R by max. 3 %.

R depends on the angle Θ_{JM} between the current density J and the magnetization M: If J and M are parallel (i.e., $\Theta_{JM}=0$), the

maximum electrical resistance $R_{||}$ results; if J and M are perpendicular to each other ($\Theta_{JM}=90°$), the minimum value $R_{\perp}$ is obtained. The term $\sin\Theta_{JM}$ can be expressed for small angles Θ_{JM} by the ratio of the applied magnetic field H_y in the hard direction and the anisotropy field strength H_K, where H_K depends on the geometry of the AMR layer – more precisely on the ratio of width b to thickness d. The following relationship applies to R (Figure 31b, [6]):

$$R(\Theta_{JM}) = R_{||} - \Delta R_{max}(\sin\Theta_{JM})^2.$$

The following applies to small Θ_{JM}:

$$\sin\Theta_{JM} \approx \frac{H_y}{H_K}.$$

$R_{||}$: Resistance for $\Theta_{JM} = 0°$;
$R_{\perp}$: Resistance for $\Theta_{JM} = 90°$;
$\Delta R_{max} = R_{||} - R_{\perp}$: Maximum achievable absolute resistance change.

Quantum physics provides a physical explanation of this effect: Changes in the electric conductivity are obtained depending on the spin orientation.

The anisotropy field strength H_K plays an important role for the sensitivity $S = dR/dH$; the wider the AMR layer, the steeper the characteristic curve and hence the sensitivity S.

It follows from Figure 31b that for small magnetic fields H_y the sensitivity S is very low; the operating point should therefore be brought into a zone with linear behavior. Above all two methods are known for this purpose: Linearization with an additional (constant) magnetic field H_{y0} (in the direction of the hard axis in Figure 31a), which shifts the characteristic curve in Figure 32 to the left into the steep zone; or the frequently used barber-pole structure in Figure 33. Here the current-density vector J is rotated through approximately 45° to obtain a linear characteristic curve for small magnetic fields H_y.

Figure 31: AMR sensor element
a) Design,
b) Characteristic curve,
1 Sensor element,
2 Contact layer,
3 Analytical calculation acc. to
$R = R_{||} - \Delta R_{max}(\sin\Theta_{JM})^2$
for $-H_K \le H_y \le H_K$ and
$R = R_{||} - \Delta R_{max}$ outside this range.
4 Real characteristic curve.
l Length of sensor element,
b Width of sensor element,
H Magnetic field,
M Magnetization,
J Current density,
R Resistance,
Θ_{JM} Angle between J and M.

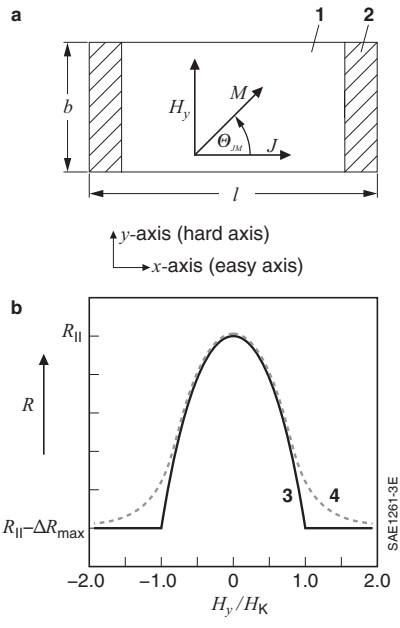
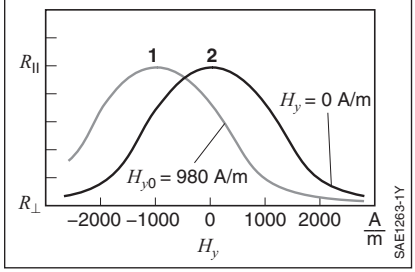


Figure 32: Linearization of the AMR characteristic curve
1 Characteristic curve with additional magnetic field H_{y0} (measured with $H_{y0} = 980$ A/m),
2 Characteristic curve without linearization.
H Magnetic field,
R Resistance.



However, it is to be noted according to Figure 34 that the orientation of the easy axis may not be changed since "rotating the domains around an axis" through 180° leads to a mirroring of the characteristic curve with regard to $H_y = 0$ ("butterfly curve"). This is as a rule effected by an additional constant magnetic field in the "easy direction" (here: x-direction), but which reduces the sensitivity and also the absolute resistance and hence also the absolute resistance change since the maximum relative change of R is always approximately 3 %. Rotating the domains around an axis results in a change of the characteristic curve during linearization with a constant magnetic field H_{y0}.

The use of AMR sensors is favored by the higher magnetic-field sensitivity obtained than with silicon Hall-effect sensors. On the down side, AMR sensors are not compatible with semiconductor processes and must therefore be designed separately. Because of the mirror-symmetric characteristic curve around zero AMR angle sensors can only resolve an angular range of 180°.

Figure 35 shows a typical circuit of four AMR elements with barber-pole structure; because of the opposing orientation of the highly conductive strips in two AMR elements respectively a piecewise linear characteristic curve is achieved, as can be seen by reference to Figure 35b.

Figure 33: AMR sensor element with barber-pole structure

1 Highly conductive layer (barber pole).
H Magnetic field,
M Magnetization,
J Current density, supply current,
Θ Angle.

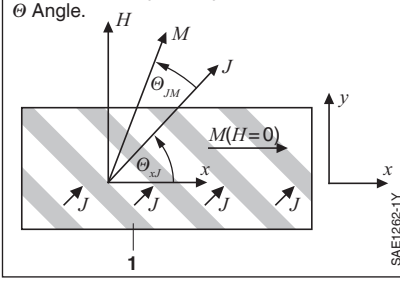


Figure 34: Linearization of the AMR characteristic curve with barber-pole structure

1 Characteristic curve 1:
Easy axis in negative x-direction,
2 Characteristic curve 2. 1:
Easy axis in positive x-direction.
H Magnetic field,
R Resistance.

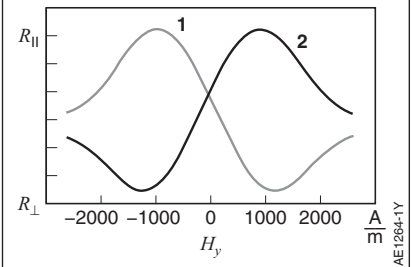
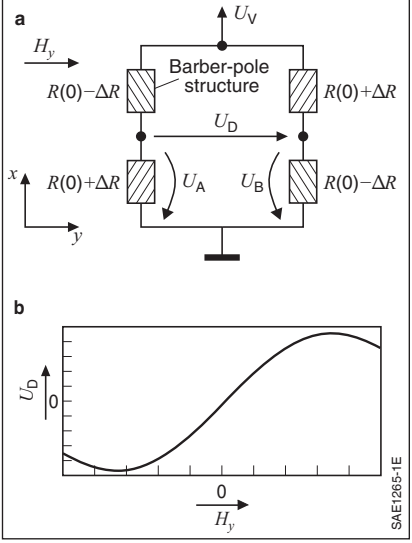


Figure 35: Wheatstone bridge with four AMR sensor elements with barber-pole structure

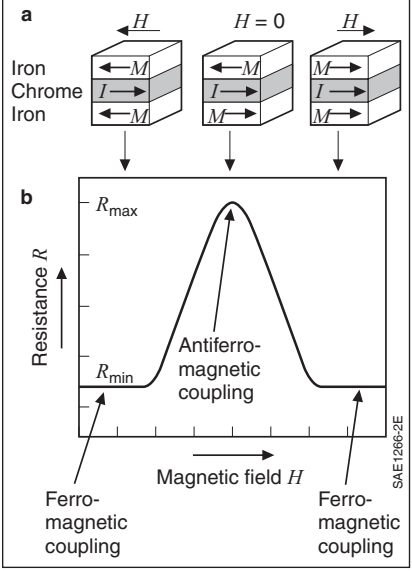
Barber-pole structure of the AMR elements is identical crosswise.
a) Design,
b) Characteristic curve.
H Magnetic field,
U_V Supply voltage,
U_A Voltage at sensor element A,
U_B Voltage at sensor element B,
U_D Bridge voltage,
ΔR Resistance change.



Giant magnetoresistive effect

In the case of the giant magnetoresistive effect (GMR effect) the electrical resistance R changes with the orientation of the magnetization M of two closely neighboring ferromagnetic layers (Figure 36, e.g., iron or cobalt) which are separated by a very thin non-magnetic intermediate layer (e.g., chrome). As with the AMR effect, only applied magnetic fields in the layer plane are relevant with regard to the GMR effect. The spin-dependent scattering of the electrons in one of the magnetic layers brings about, with a parallel orientation of the magnetization M (i.e., ferromagnetic coupling), the minimum electrical resistance R_{min} in the two layers; with an antiparallel orientation of the two magnetizations M (thus antiferromagnetic coupling) the maximum value R_{max} is the result. Relative resistance changes of approximately 6 % are obtained when a number of layer sequences are used.

Figure 36: Characteristic curve and basic design of a GMR sensor element
a) Basic design,
b) Characteristic curve.
M Magnetization,
I Current,
R Resistance,
H Magnetic field.

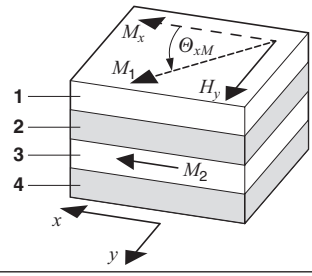


The current I flows in GMR sensors preferably in the layer plane of the magnetization M (CIP-GMR, Current-in-plane-GMR). However, a current flow perpendicular to the layer plane (CPP-GMR, Current-perpendicular-to-plane-GMR) is also possible. Figure 36b shows the bell shape of R (similar to the AMR effect) as a function of an applied magnetic field H.

The important aspect here is the minimal thickness of the electrically conductive and non-magnetic intermediate layer (chrome in Figure 36 or "spacer" in Figure 37; copper is also used), which is in the region of below one nanometer.

GMR sensor elements are frequently used as spin valves in order to generate a sufficient and at least piecewise linear signal in the case of small magnetic fields H. The basic design is shown in Figure 37: Here the magnetization M_2 of a ferromagnetic layer is retained (in the "pinned layer" with antiferromagnetic coupling) so that only the ("rotatable") magnetization M_1 of the soft-magnetic layer follows an external magnetic field H_y; M_x in Figure 37 corresponds to the orientation of M_1 without an external magnetic field H_y. The resulting characteristic curve is shown in Figure 38; this characteristic curve shows for small magnetic fields H_y an at least piecewise linear shape – however hysteresis oc-

Figure 37: Design of a spin valve
1 Soft-magnetic layer,
rotatable magnetization M_1,
2 "Spacer", metallic non-magnetic individual layer,
3 Ferromagnetic layer,
fixed magnetization M_2,
4 Antiferromagnetic coupling layer.
M Magnetization,
H Magnetic field,
Θ_{xM} Angle between x and M.



curs with higher magnetic fields H_y. The maximum relative change of R for GMR spin-valve sensors is in the range of approximately 4%.

GMR sensor elements are frequently designed in a Wheatstone bridge circuit – i.e., as a full bridge.

The advantages of GMR sensors in comparison with AMR sensors lie in their higher sensitivity S (= dR/dH) and smaller space requirement, i.e., more signal per surface area. Furthermore, a higher resolution can be achieved on account of the smaller GMR sensor elements. All this permits the use of smaller magnets for angle sensors and smaller gear wheels for example for rpm sensors.

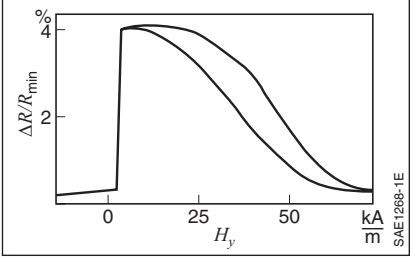
In addition, angle sensors with a measurement range of 360° are possible with GMR sensors, while only 180° can be achieved with AMR sensors. The reason for this is that with GMR sensors the orientation of the pinned magnetization can be arranged differently for the individual GMR sensor elements.

Tunnel magnetoresistive effect

In the case of the tunnel magnetoresistive effect (TMR effect), as with the GMR effect, the electrical resistance R changes with the orientation of the magnetization M of two closely neighboring ferromagnetic layers. However, with TMR sensors, the current I does not, as with customary CIP-GMR sensors, flow parallel (compare here Figure 36), but instead flows due to the tunnel effect, perpendicularly to the electrically insulating and non-magnetic intermediate layer. As with the AMR and GMR effects, only applied magnetic fields in the layer plane are relevant with regard to the TMR effect.

Much smaller and even more sensitive sensors compared with GMR and AMR sensors can be designed with TMR technology; this is due to the much higher electrical resistance R for comparative relative resistance changes. The electrical power demand is reduced here in accordance with $P = U^2/R$.

Figure 38: Example of a characteristic curve of a spin-valve GMR sensor element
H Magnetic field,
R Resistance.



References

[1] H. Bernstein: Elektrotechnik/Elektronik für Maschinenbauer – Einfach und praxisgerecht. 3rd Ed., Verlag Springer Vieweg, 2018.

[2] D. Zastrow: Elektrotechnik – Ein Grundlagenlehrbuch. 20th Ed., Verlag Springer Vieweg, 2018.

[3] E. Philippow: Grundlagen der Elektrotechnik. Verlag Technik, Huss-Medien, 10th Ed., 2000.

[4] D. Meschede: Gerthsen Physik. 25th Ed., Verlag Springer Spektrum, 2015.

[5] http://www.uni-magdeburg.de/exph/messtechnik1/Parameter_Thermoelemente.pdf.

[6] E. Kneller: Ferromagnetismus. 1st Ed., 1962; Reprint 2012, Springer-Verlag.

Electronics

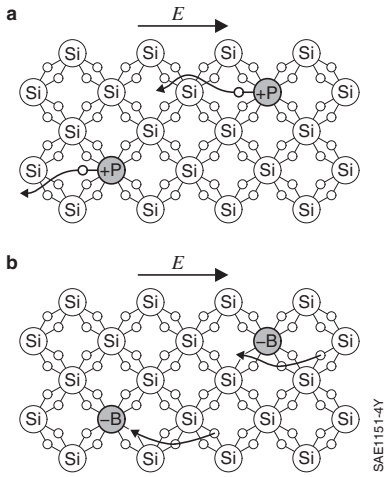
Basic principles of semiconductor technology

Electric conductivity in solid bodies

An individual material's capacity for conducting electricity is determined by the number and mobility of the free charge carriers which it contains. The disparities in the electric conductivities displayed by various solid bodies at room temperature extend through a range defined by 10 to the 24th power. Accordingly, materials are divided into three electrical groups.

Conductors (metals)
All solid bodies contain approximately 10^{22} atoms per cubic centimeter; they are held together by electrical forces. In metals, the number of free, i.e. non-attached, charge carriers is extremely high (one to two free electrons per atom). The free carriers are characterized by moderate mobility. The electrical conductivity of metals (e.g. silver, copper, aluminum) is high. In good conductors, this amount to approx. 10^6 S/cm.

Figure 1: Doped silicon
a) *n*-doped silicon,
b) *p*-doped silicon.
o Electron, Si Silicon, P Phosphorus, B Boron, *E* Electric field.
The curved arrows indicate the direction of motion of the electrons when electric field *E* is applied.



Non-conductors (insulators)
The number of free charge carriers in insulators (ex. aluminum oxide, Teflon, fused quartz) is practically zero. Accordingly, the electric conductivity is negligible. The conductivity of good insulators is approximately 10^{-18} S/cm.

Semiconductors
The amount of electrical conductivity in semiconductors (ex. germanium, silicon, gallium arsenide) is between that of metals and insulators. It is – unlike the conductivity of metals and insulators – greatly dependent on the following factors:
– the pressure affects the mobility of charge carriers,
– the temperature influences the number and mobility of charge carriers,
– the action of light also influences the number of charge carriers,
– the presence of additives also determines among others the number and type of charge carriers.

Because they are dependent on the above factors, semiconductors are also suitable for use as pressure, temperature and light sensors.

Doping of semiconductors

Doping, i.e. the controlled addition of electrically active foreign substances to the base material, makes it possible to define and localize the semiconductor's conductivity. This procedure forms the basis of semiconductor devices. The producible and also adjustable electric conductivity of silicon that can be reproduced by doping is 10^4 to 10^{-2} S/cm.

Electric conductivity of semiconductors
The following discussion focuses on the silicon-based semiconductor. In its solid state, silicon assumes the form of a crystal lattice with four equidistant contiguous atoms. Each silicon atom has four outer electrons, with two shared electrons form-

ing the bond with the contiguous atoms. In this ideal state, silicon has no free charge carriers; thus it is not conductive. The situation changes dramatically with the addition of appropriate additives and the application of energy.

Doping will be explained here by reference to a clear model presentation. However, it is important to bear in mind that not all effects can be explained by reference to this model [2].

n Doping

As only four electrons are required for bonding in a silicon lattice (Figure 1a), the introduction of foreign atoms with five outer electrons (e.g. phosphorus) results in the presence of free electrons. Thus, each additional phosphorus atom will provide a free, negatively charged electron, where a single positively charged phosphorus nucleus remains. The silicon is transformed into an n conductor (n-type silicon), since there is an excess of negative charges (electrons). In response to an externally applied voltage, an electric field E is generated, which gives the mobile charge carriers a preferred direction for movement (Figure 1).

p Doping

The introduction of foreign atoms with three outer electrons (e.g. boron) produces electron holes which result from the fact that the boron atom has one electron less for complete bonding in the silicon lattice (Figure 1b). This hole indicates a missing electron. Holes remain in motion within the silicon; they are filled by electrons that leave behind their own hole. In an electric field, they travel in the opposite direction of the electrons. The holes exhibit the properties of a free positive charge carrier. Thus, every additional boron atom provides a free, positively charged electron hole (positive hole). The silicon is transformed into a p conductor and is called a p-type silicon.

Intrinsic conduction

Heat and light also generate free mobile charge carriers in undoped silicon; the resulting electron-hole pairs produce intrinsic conductivity in the semiconductor material. This conductivity is generally modest in comparison with that achieved by doping. Increases in temperature induce an exponential rise in the number of electron-hole pairs, ultimately obviating the electrical differences between the p and n regions produced by doping. This phenomenon defines the maximum operating temperature to which semiconductor devices may be subjected: This is for germanium 90 to 100 °C, for silicon 150 to 200 °C, and for gallium arsenide 300 to 350 °C.

A small number of opposite-polarity charge carriers is always present in both n-type and p-type semiconductors. These minority charges exert a considerable influence on the operating characteristics of virtually all semiconductor devices.

pn junction in the semiconductor

The area of transition between a p-type zone and an n-type zone within the same semiconductor crystal is referred to as the pn junction. The properties of this area exercise a major influence on the operating properties of most semiconductor devices.

pn junction without external voltage

The p-type zone has numerous holes, and the n-type zone has only very few; there is a very large number of electrons in the n-type zone, while in the p-type zone there is an extremely small number. Each type of mobile charge carrier tends to move across the concentration gradient, diffusing into the other zone (Figure 2b).

The diffusion of holes into the n-type zone gives the p-type zone in the pn junction area a negative charge, since the negatively charged core atoms (i.e. boron atoms) remain stationary. Electron depletion results in the n-type region being positively charged, since here this is an excess of stationary, positively charged atomic residues (e.g. phosphorus). The result is an electrical potential between the p- and n-type zones (diffusion potential U_D). This potential opposes the respective migration tendencies of the charge carriers. This ultimately brings the exchange of holes and electrons to a halt. The potential current U_D created on account of diffusion cannot be directly measured from the outside and in silicon is typically just 0.6 V.

A poorly conducting zone deficient in mobile charge carriers is thus produced at the *pn* junction: This zone is called the space-charge region or depletion layer. This zone has an electric field whose strength is also dependent on the externally applied voltage.

pn junction with external voltage
Now the conditions at a diode will be explained because a *pn* junction corresponds to the structure of a diode; here the anode is situated at the *p*-doped silicon, while the cathode is situated at the *n*-doped zone.

The application of voltage U in the reverse direction (negative pole at the *p*-type zone and positive pole at the *n*-type zone) extends the space-charge region (Figure 2c). Consequently, the flow of current I is blocked except for a minimal residual current (reverse current) which stems from the minority charge carriers. The voltage U then drops within the space-charge region; this region is therefore subject to a high electric field strength.

Breakdown voltage is the level of reverse-direction voltage beyond which a minimal increase in voltage will suffice to produce a sharp rise in reverse current (Figure 3). This effect can be explained as follows: Electrons which reach the space-charge region are greatly accelerated on account of the high field strength. In this way, they can for their part generate free charge carriers as a result of impacts; this is also known as collision ionization. This brings about a dramatic rise in current and results in avalanche breakdown. In addition to avalanche breakdown, there is also Zener breakdown, which is based on the tunnel effect. The breakdown of a *pn* junction can destroy it and for this reason is often not wanted. In many cases, however, breakdown is wanted. Avalanche breakdown and Zener breakdown only occur when the diode is operated in the reverse direction.

The application of a voltage U in the forward direction (positive pole at the *p*-type zone and negative pole at the *n*-type zone) reduces the space-charge region (Figure 2d). Charge carriers permeate the *pn* junction, resulting in a large flow of current in the forward direction (Figure 3), as the space-charge region no longer has a significant resistance. Only the bulk resistance, i.e. the ohmic resistance of the doped layers, is effective. The current I increases exponentially as a function of U. It is, however, important to be aware of "thermal breakdown", at which point the semiconductor can be destroyed on account of the high level of heat. This can occur, for example, if the diode is operated in the forward direction with an unacceptably high current.

Figure 2: *pn* junction in a diode
a) Diode circuit symbol,
b) *pn* junction without external voltage,
c) *pn* junction in reverse direction,
d) *pn* junction in forward direction.
U Applied voltage (diode voltage),
I Diode current.
⊕ Positively charged atomic residues,
⊖ Negatively charged atomic residues.

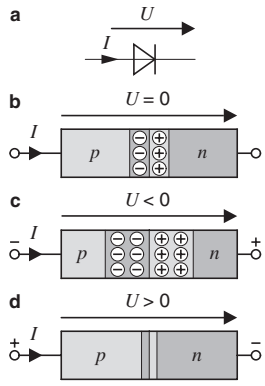
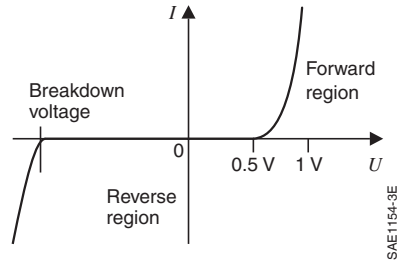


Figure 3: Characteristic curve of a silicon diode
U Applied voltage (diode voltage),
I Diode current.



Discrete semiconductor devices

The properties of the *pn* junction and the combination of several *pn* junctions in a single semiconductor-crystal chip provide the basis for a steadily increasing array of inexpensive, reliable, rugged, compact semiconductor devices. A single *pn* junction forms a diode, two *pn* junctions are used for transistors. The planar technique makes it possible to combine numerous operating elements on a single chip to form the important component group known as integrated semiconductor circuits. Semiconductor chips measure no more than several square millimeters and are usually installed in standardized housings (metal, ceramic or plastic).

Diodes

A diode is a semiconductor device incorporating a single *pn* junction. Its specific properties are determined by the distribution pattern of the dopant in the crystal. Diodes which conduct currents in excess of 1 A in the forward direction are frequently referred to as power diodes.

Rectifier diode

The rectifier diode acts as a form of current valve; it is, therefore, ideally suited to rectifying alternating current. The forward current can be approximately 10^7 times higher than the current in the reverse direction (reverse current, Figure 3). As the temperature rises, the power of this current increases significantly.

Diodes can also be used as polarity-reversal protection diodes in order to prevent a current flow in the event of an incorrect voltage connection. It is also customary to use them as free-wheeling diodes.

Rectifier diode for high reverse voltage

In the case of a rectifier, the voltage drops over the space-charge region. Because this region is generally only a few micrometers in size, the high reverse currents produce a high electric field strength and the free electrons can accelerate greatly. Accelerated electrons can cause the semiconductor to be destroyed (avalanche breakdown). To prevent this, it has proved useful to integrate an intrinsic layer between the *p*- and *n*-layers because this layer only contains few free electrons and thereby reduces the danger of a breakdown.

Switching diode

A switching diode is generally employed for rapid switching between high and low impedances. More rapid switching response can be achieved by diffusing gold into the material; this promotes the recombination of electrons and holes. During switching the space-charge region must be recharged accordingly; during switching-on from the blocking to the conducting state the initially charge-free space-charge region is filled with charges, which results in the forward recovery time. Accordingly, when the diode passes into the blocking state the (excess) charge carriers located in the space-charge region must be removed – the time needed for this is called the backward recovery time. For bipolar transistors the saturation (U_{CE} < 0.2 V) is therefore critical when rapid switching is required. The storage charge or the maximum reverse current is often also specified here.

A distinction is made in the losses between forward power losses (forward voltage times current) and switching power losses; at lower frequencies up to a few 100 Hz the forward power losses dominate while at higher frequencies the switching power losses increasingly play role.

Power diodes

When compared with switching diodes, power diodes are designed for higher currents (> 1 A) and high reverse voltages (up to over 1 kV). These demands arise from the conditions in power electronics where high current intensities and high (reverse)voltages can occur with low forward power losses being desired. Power diodes therefore frequently have – similarly to rectifier diodes for high reverse voltages – a lightly doped "middle layer" between the *p*- and *n*-doped regions – i.e. a pin structure: *p*-doped, undoped or intrinsic, *n*-doped. The space-charge region can further expand in the lightly doped region and thus a higher (reverse) voltage

can be achieved without the dopings having to be reduced in the p and n regions.

Aside from pure silicon diodes, Schottky diodes based on silicon carbide (SiC) are also used here; these diodes exhibit a forward voltage of 0.8 V and a maximum reverse voltage of barely 2 kV. Furthermore, SiC diodes can also be operated at higher temperatures (up to 200 °C).

Free-wheeling diodes
High voltages can occur during the rapid switching of currents by inductance since the induced voltage is proportional to dI/dt. Frequently free-wheeling, or flyback, diodes are used to prevent damage to components such as power transistors (e.g. MOSFET or IGBT). In this case the current flow is not abruptly interrupted, but can continue to flow via the free-wheeling diodes so that high induced voltages are avoided.

Zener diode
A Zener diode is a semiconductor diode which, once a specific initial level of reverse voltage is reached, responds to further increases of reverse voltage with a sharp rise in current flow. This phenomenon is a result of a Zener and/or avalanche breakdown. Zener diodes are designed for continuous operation in this breakdown range. These are frequently used to provide a constant voltage or reference voltage.

Variable-capacitance diode (varactor)
The space-charge region at the pn junction functions as a capacitor; the dielectric element is represented by the semiconductor material in which no charge carriers are present. Increasing the applied voltage extends the depletion layer and reduces the capacitance, while reducing the voltage increases the capacitance.

Schottky barrier diode (Schottky diode)
A Schottky diode contains a metal-to-semiconductor junction. As the electrons move more freely from the n-type silicon into the metal layer than in the opposite direction, an electron-depleted layer is created in the semiconductor material; this is the Schottky barrier layer. Charges are carried exclusively by the electrons, a factor which results in extremely rapid switching, as the minority carriers do not perform any charge storage function (charge-carrier surplus in the space-charge region is reduced). The forward voltage and thus the voltage drop is at approximately 0.3 V smaller in Schottky diodes than in silicon diodes (approximately 0.6 V). Because of the rapid switching and the comparably low losses, Schottky diodes are used in high-frequency circuits down into the microwave range.

Solar cell
The photovoltaic effect is applied to convert light energy directly into electrical energy. Solar cells, consisting largely of semiconductor materials, are the basic elements of photovoltaic technology. Exposure to light can cause the formation of free charge carriers (electron-hole pairs) in the semiconductor. If the semiconductor incorporates a pn junction, the free charge carriers separate in its electric field before proceeding to the metal contacts on the semiconductor's surface. Depending on the semiconductor material used, a DC voltage (photovoltage) ranging from 0.5 to 1.2 V is generated between the contacts. This occurs only when the light quanta have at least the energy required to create an electron-hole pair; the energy is proportional to the frequency and inversely proportional to the wavelength of the light. The theoretical efficiency of crystalline silicon solar cells is approximately 30 %.

Photodiode
The photovoltaic effect is utilized in a photodiode. The pn junction is operated in the reverse direction. Incident light creates additional free electrons and holes. These increase the reverse current (photovoltaic current) in direct proportion to the intensity of the light. The photodiode is thus, in principle, very similar to the solar cell.

Light-emitting diode
A light-emitting diode or LED is an electroluminescent lamp. which consists of a semiconductor element with pn junction. The charge carriers (free electrons and holes) recombine during operation in forward direction. The amount of energy

released in this process is converted into electromagnetic radiation energy.

Depending on the choice of semiconductor and its doping, the LED emits in a limited spectral range. Frequently used semiconductor materials are: gallium arsenide (infrared), gallium arsenide phosphide (red to yellow), gallium phosphide (green), and indium gallium nitride (blue). In order to generate white light, either a combination of three LEDs with the primary colors Red, Green and Blue is used, or a fluorescent dye is excited with a blue or ultraviolet-emitting LED. The maximum luminous efficiency of present-day LEDs is over 300 lm/W – the maximum possible value is approximately 350 lm/W, where only the spectrum that can be detected by the human eye is taken into account here (photometric quantity "luminous flux" in lumens or abbreviated as "lm").

OLED

OLED stands for "Organic Light-Emitting Diode". An OLED is a very thin, two-dimensional light source which emits soft and for the most part non-dazzling light – sharp cast shadows do not occur. OLEDs are in a good approximation Lambertian emitters, i.e. they always appear with uniform brightness regardless of the viewing angle. The small geometric dimensions are also to be noted; thus, OLEDs are less than 2 mm thick – including glass substrate and encapsulation. A further particularity is the flexibility of the extremely flat design, which creates further possible applications in addition to the light properties with high color rendering.

OLEDs - as non-dazzling large-area emitters with a broad spectrum - are particularly suitable for interior applications in automobiles. Constant-current operation suggests itself as the form of activation, enabling light changes on account of aging, temperature or even by production tolerances to be minimized. OLEDs can be dimmed for example by pulse-width modulation.

During the manufacture of OLEDs a number of thin organic semiconductor layers are placed in succession on the conductive substrate (formation of a stack); these layers can form polymeric substances or "small-molecule materials" as an amorphous layer.

Further conductive and transparent electrodes, which cover the organic layers on both sides, are then used. Oxides such as indium tin oxide (ITO) are used as the material for these electrodes; on account of the non-ideal conductivity there is a voltage drop in the active layers from the outside inwards, which results in a corresponding reduction of the radiation intensity from the outside inwards. A remedy can be provided by an auxiliary structure made of more conductive material. However, the homogeneity of the radiation intensity can be improved by the multi-layer design (stacking).

The maximum luminous efficiency for OLEDs is today approximately 100 lm/W – i.e. still significantly below that of LEDs. However, the service life must also be noted; this decreases with higher operating current – and consequently higher light output.

Laser LED

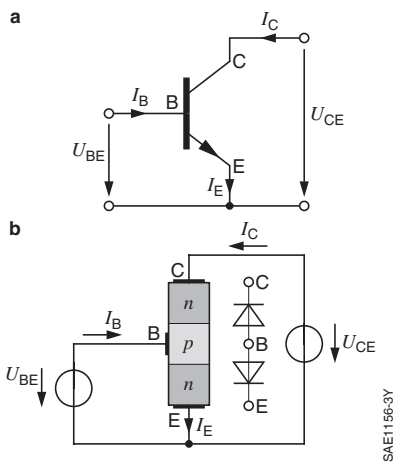
New lighting systems increasingly use laser LEDs – semiconductor-based laser light sources. These are small spotlights with very high luminance (a factor of approximately 4 compared with LEDs). It is to be noted here that the initially generated blue laser beam ($\lambda \approx 450$ nm) does not leave the headlamp, but is converted with the aid of luminescent materials (ceramic high-power materials) into visible white light with a color temperature of around 5,500 K. A light cone with a very high range (up to 600 m) can then be achieved with very small optical lenses. This makes possible small headlamps with high light output. The performance of a laser LED is decisively influenced by the properties of these luminescent materials and their optical properties. Hybrid light concepts are frequently used at present so that LEDs and laser LEDs are used for the driving lights and for the high-beam lights respectively.

Bipolar transistors

Two contiguous *pn* junctions produce the transistor effect, a feature employed in the design of components used to amplify electrical signals and to assume switching duties. Bipolar transistors consist of three zones of varying conductivity, the configuration being either *pnp* or *npn*. The zones (and their terminals) are called emitter E, base B, and collector C (Figure 4).

There are different transistor classifications, depending on the fields of application: small-signal transistors (power dissipation up to 1 watt), power transistors, switching transistors, low-frequency transistors, high-frequency transistors, microwave transistors, and phototransistors. They are termed bipolar because charge carriers of both polarities (holes and electrons) are involved in the transistor effect.

Figure 4: *npn* transistor
a) Diagram,
b) Structure.
E Emitter,
B Base,
C Collector.
U_{BE} Base-emitter voltage,
U_{CE} Collector-emitter voltage.
I_B Base current,
I_C Collector current,
I_E Emitter current.



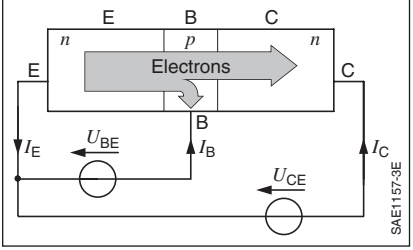
Operation of a bipolar transistor

The operation of a bipolar transistor is explained here using an *npn* transistor as the example (Figure 5). The *pnp* transistor is obtained similarly by switching the *n*- and *p*-doped zones. The base-emitter junction is forward-biased; this is shown in Figure 4b as a diode between base B and emitter E. In this way, with sufficient voltage U_{BE}, electrons are injected into the base zone and base current flows.

The base-collector junction is reverse-biased; this is shown in Figure 4b as a diode between base B and collector C. This creates a space-charge region at the *pn* junction between base and collector with a high electric field.

Because of the reverse-biased diode between the base and the emitter, a high current consisting of electrons flows from the emitter to the base. Here, however, only a small fraction can recombine with the (much fewer) holes and flow as base current I_B out of the base terminal; what must be borne in mind is that the technical current direction – i.e. the direction of motion of the positive charge carriers – is specified in Figure 4. The much greater amount of electrons injected into the base diffuses through the base zone to the base-collector junction and then flows as collector current I_C to the collector (Figure 5). Because the base-collector diode

Figure 5: Operation of an *npn* transistor
E Emitter,
B Base,
C Collector.
U_{BE} Base-emitter voltage,
U_{CE} Collector-emitter voltage,
I_B Base current,
I_C Collector current,
I_E Emitter current.



is operated in the reverse direction and a space-charge region is predominant, virtually all (approximately 99 %) of the electrons flowing from the emitter are "drawn off" through the strong electric field in the space-charge region from the collector. In this case, a linear connection applies approximately between the collector current I_C and the base current I_B:

$$I_C = B\, I_B$$

where B is the current gain, which is generally between 100 and 800. In the bipolar transistor, the following relation for the emitter current I_E also applies (cf. Figure 4 and Figure 5):

$$I_E = I_B + I_C\,.$$

Assuming that I_B on account of the current gain B is much smaller than I_C, it then follows:

$$I_E \approx I_C\,.$$

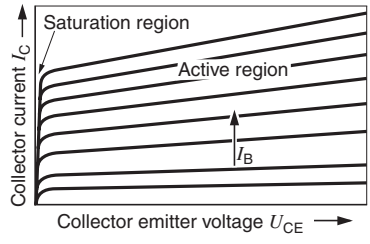
The very thin (and relatively low-doped) base represents a barrier that can be adjusted by means of the base-emitter voltage U_{BE} from the charge-carrier flow from the emitter to the collector. With a small change of U_{BE} and the base current I_B, it is possible to control a greater change of the collector current I_C and the collector-emitter voltage U_{CE}. Small changes in the base current I_B thus bring about large changes in the emitter-collector current I_C. The *npn* transistor is a bipolar, current-controlled semiconductor amplifier. In all, there is a power amplification.

The output characteristic curve for an *npn* transistor is shown in Figure 6. From the saturation voltage of approximately 0.2 V for U_{CE}, the collector current I_C is virtually only dependent on the base current I_B as the parameter; this region is termed the "active region": U_{CE} here has virtually no influence on I_C and the following applies:

$$I_C = B\, I_B.$$

The region below the saturation voltage is called the "saturation region". In this region, I_C increases sharply with U_{CE}.

Figure 6: Output characteristic curve of an *npn* transistor
U_{CE} Collector-emitter voltage,
I_C Collector current,
I_B Base current as parameter of characteristic curve.



Darlington transistor

The Darlington arrangement of two bipolar transistors is often used to achieve a higher current amplification; the interconnection of the two transistors is also called a Darlington transistor (Figure 7). Transistor T_1 (as emitter follower) switches transistor T_2 so that in total a very high current amplification β (up to 50,000) is achieved as the product of the two current amplifications of the two transistors T_1 and T_2: $\beta = \beta_1 \cdot \beta_2$ ($\approx I_{C2}/I_{B1}$). Compared with power bipolar transistors ($\beta \approx 5\ldots10$), the current amplification is much higher and consequently the control current is much lower.

Darlington transistors are used when a voltage (which is not to be subjected to load) connects as high a load as possible (i.e. current). On account of the phase displacement between the input (control current I_{B1} to the base of T_1) and the output (collector current I_{C2} in T_2), the Darlington transistor is only suitable for relatively low frequencies – this circuit is therefore not used in high-frequency circuits. The main reason for this is the time required for the switching of T_2 through the removal of the excess charges in the base of T_2 via resistance R; if R is selected as too small, the total current amplification β is reduced.

Another drawback is the voltage at the input of 1.2…1.4 V, which is twice as high compared with simple bipolar transistors. The forward voltage U_{CE2} at T_2 increases by U_{BE2} to approximately 0.9 V (compared with the saturation voltage of 0.2 V for simple bipolar transistors – here U_{CE1}) or approximately 2 V for power types. As a result the losses are increased (proportionally $U_{CE2} \cdot I_{C2}$).

Diode D_3 is a free-wheeling diode for inductive loads in order to avoid high voltages between the emitter and collector of T_2.

IGBT

The insulated-gate bipolar transistor (IGBT) is an important switching element in the middle power range for currents up to 3 kA and voltages up to 3 kV. The IGBT is characterized by relatively low switching power losses and high efficiency.

The IGBT combines the advantages of a MOSFET (T_1 in Figure 8, see Field-effect transistors) with low-power, voltage-controlled activation and of a power bipolar transistor T_2 with a relatively low forward voltage at the output: In Figure 8 voltage U_{GE} is used at the MOS transistor T_1 to switch on the *pnp* bipolar transistor T_2, which then directs the output current I_C with relatively low forward voltage U_{CE}; the saturation voltage U_{CESAT} is in the range of 2…3 V.

Due to the curved decay of the current during switch-off on account of the light doping, the switching frequency is limited – as a rule 20 kHz to max. 100 kHz for smaller power levels.

Field-effect transistors

In a field-effect transistor (FET), control of the current flow in a conductive path is essentially exercised by an electric field. The field, in turn, is generated by a voltage applied at a gate (Figure 9). Field-effect transistors differ from their bipolar counterparts in that they utilize only a single type of charge carrier (either electrons or holes), giving rise to the alternate designation "unipolar transistors". These are subdivided into the following categorizations: junction-gate field-effect transistors (junction FET, JFET) and insulated-gate field-effect transistors, particularly MOS field-effect transistors (MOSFET or MOS transistors).

MOS field-effect transistors are well suited for application in high-integration circuits. Power field-effect transistors represent a genuine alternative to bipolar power transistors in many applications.

The advantages of a bipolar transistor and of a field-effect transistor are utilized in power electronics in "insulated-gate bipolar transistors" (IGBT), which exhibit a low volume resistance (for small losses) and comparatively low triggering power.

Figure 7: Circuit of a Darlington transistor
T Bipolar transistors, R Resistor, E Emitter, B Base, C Collector, D Diode.
I_C Collector current,
U_{BE} Base-emitter voltage,
U_{CE} Collector-emitter voltage.

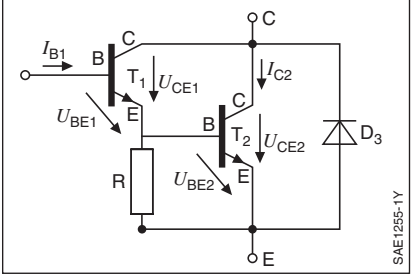
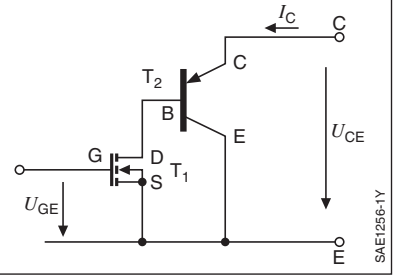


Figure 8: Circuit of an IGBT
T_1 Field-effect transistor,
T_2 Bipolar transistor.
G Gate,
S Source,
D Drain.
U_{GE} Gate-emitter voltage,
I_C Collector current.

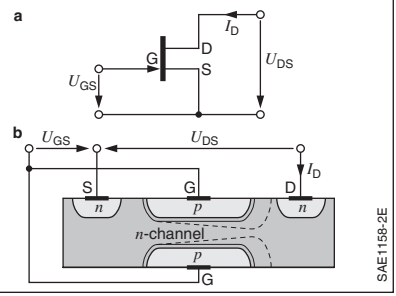


Operation of a junction FET
The operation of a junction-gate field-effect transistor is explained by reference to the n channel type (Figure 9). The terminals of a field-effect transistor are referred to as gate (G), source (S) and drain (D).

Positive direct voltage U_{DS} is applied at the ends of an n-type crystal. Electrons flow through the channel from the source to the drain. The width of the channel is defined by two laterally diffused p-type zones and by the negative gate-source voltage U_{GS} applied at them. The voltage U_{GS} between the control electrode (gate G) and the source terminal thus controls the current I_D between the source and the drain.

Only charge carriers of one polarity are required for field-effect transistor operation. The power necessary for controlling the current is virtually nil. Thus, the junction FET is a unipolar, voltage-controlled component. Increasing U_{GS} causes the space-charge regions to extend further into the channels, thereby constricting the channel and thus the current path (see dashed lines in Figure 9). If the voltage U_{GS} at the gate is zero, the channel between the two p-type zones is not constricted and the current I_D from the drain to the source is at its maximum.

Figure 9: Junction-gate field-effect transistor with *n* channel
a) Diagram,
b) Structure.
The lightly shaded region around the source and drain contacts is more heavily doped than the channel.
G Gate, S Source, D Drain.
U_{DS} Drain-source voltage,
U_{GS} Gate-source voltage,
I_D Drain current.

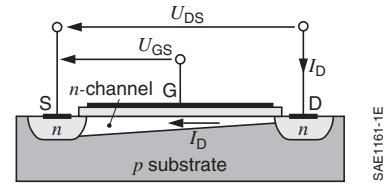


The transfer characteristic curve – i.e. I_D as a function of U_{GS} – then looks exactly the same as the characteristic curve of a self-conducting n-channel MOSFET as shown in Figure 11c.

Operation of an MOS transistor
The operation of an MOS transistor (metal-oxide semiconductor) is explained by reference to the self-blocking n-channel MOSFET (enhancement type) (Figure 10). If no voltage is applied to the gate, then no current will flow between the source and the drain: the pn junctions remain in blocking mode. The application of a positive voltage at the gate causes, on account of the electrostatic induction in the p-type zone below this gate, the holes to be displaced toward the interior of the crystal and electrons – which are always present in p-type silicon as minority charge carriers – to be pulled to the surface. A narrow n-type layer, an n channel, forms below the surface. Current can now flow between the two n-type zones (source and drain). This current consists exclusively of electrons. As the gate voltage acts through an insulating oxide layer, no stationary current flows through the gate; no power is required for the control function. It is required merely to activate and deactivate electrical power in order to recharge the gate capacity. In summary, the MOS transistor is a unipolar, voltage-controlled component.

In the case of the self-conducting n-channel MOSFET (depletion type, Figure 11a), the gate-source voltage U_{GS}

Figure 10: *n*-channel MOSFET, cross-section
S Source, G Gate, D Drain.
U_{DS} Drain-source voltage,
U_{GS} Gate-source voltage,
I_D Drain current.



is between the here negative threshold voltage U_T and zero volts (Figure 11c). At U_{GS} = 0 V, the self-conducting n-channel MOSFET has a channel below the gates for the current flow. Figure 11c shows I_D as a function of U_{GS}, where the circuit is operated in the active region with sufficient and constant U_{DS}. The transfer characteristic curve is a parabola. In contrast, the self-blocking n-channel MOSFET (Figure 11b) conducts only from the positive threshold voltage U_T^* > 0 V here (see Figure 11c). The self-blocking MOSFET is much more common that the self-conducting MOSFET.

The output characteristic curve of a self-blocking n-channel MOSFET is shown in Figure 12. The region below the knee-point voltage U_K, i.e. for $U_{DS} < U_K$ is referred to, on account of the linear characteristic curve, as the linear or ohmic region; here the MOSFET behaves like an ohmic resistance. Above the knee-point voltage U_K, i.e. for $U_{DS} > U_K$, the output current I_D is virtually uninfluenced by the drain-source voltage U_{DS}; this region is known as the cut-off region. The amount of I_D is dependent only on the gate-source voltage U_{GS}. The formula is:

$$I_D = 0.5\, K\, (U_{GS} - U_T)^2$$

where K is the proportionality number (dependent among others on technological quantities) and U_T the threshold voltage from which the transistor conducts, i.e. a channel forms (see Figure 11c).

PMOS, NMOS, CMOS transistors

As well as the n-channel MOSFET (NMOS transistor), mixing the doping produces the PMOS transistor. As the electrons in the NMOS transistor are more mobile, it operates more rapidly than the PMOS device, although the latter was the first to become available due to the fact that it is physically easier to manufacture.

It is also possible to employ complementary MOS technology to pair PMOS and NMOS transistors in a single silicon chip; the resulting devices are called complementary MOS transistors (CMOS transistors, Figure 13). The specific advantages of the CMOS transistor are extremely low power dissipation, a high degree of immunity to interference, relative insensitivity to varying supply voltages, and suitability for analog signal processing and high-integration applications [2].

Figure 11: *n*-channel MOSFET
a) Diagram of self-conducting n-channel MOSFET,
b) Diagram of self-blocking n-channel MOSFET,
c) Characteristic curves.
1 Characteristic curve of self-conducting n-channel MOSFET,
2 Characteristic curve of self-blocking n-channel MOSFET.
U_{GS} Gate-source voltage,
I_D Drain current,
U_{DS} Drain-source voltage,
U_T, U_T^* Threshold voltage.

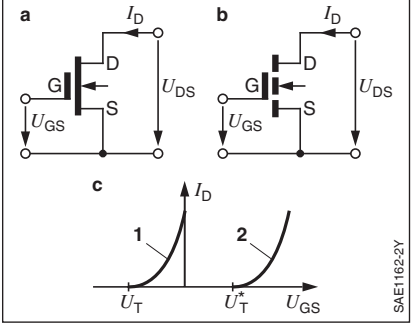
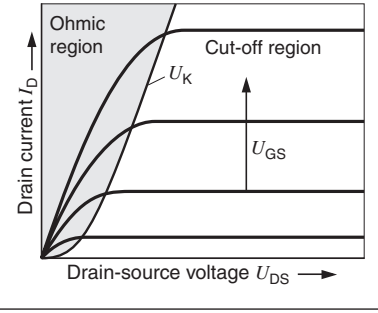


Figure 12: Output characteristic curve of a self-blocking *n*-channel MOSFET
U_{DS} Drain-source voltage,
U_{GS} Gate-source voltage,
I_D Drain current,
U_K Knee-point voltage.

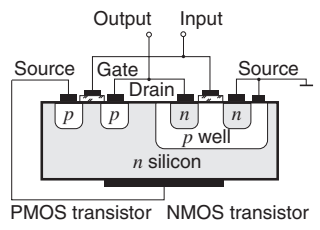


BCD hybrid technology
Integrated structures for power-electronics applications are becoming increasingly important. Such structures are realized by combining bipolar and MOS components on a single silicon chip, thereby utilizing the advantages of both technologies. BCD hybrid technology is a significant manufacturing process in automotive electronics and also facilitates the manufacture of MOS power components (DMOS). This technology is a combination of bipolar, CMOS and DMOS technologies [2].

Operational amplifiers
Areas of application of operational amplifiers
The name "operational amplifier" (OPA) comes from analog computing technology and denotes an (almost) ideal amplifier. Because of its properties, it was used particularly in analog computers to solve nonlinear differential equations, i.e. as a summator, integrator and differentiator. The rapid development of digital electronics saw analog computers being increasingly driven from the market with the result that today analog computers do not play a role at all.

By integrating them in microelectronic circuits, it is today possible to offer such operational amplifiers at a very economical price on the market so that many amplifier applications can be put into effect. To achieve the desired properties, operational amplifiers in their integrated form contain some (depending on the requirement 10 to 250) transistors, the number of which however is only of subordinate importance with regard to integration.

Figure 13: CMOS inverter composed from PMOS and NMOS technology



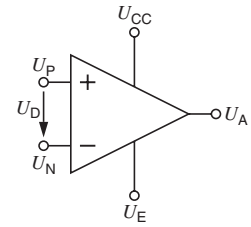
Starting out from a "normal" operational amplifier with voltage input and voltage output (VV operational amplifier), first the behavior of an ideal operational amplifier will be described and its use demonstrated. Then the real – i.e. actual – properties will be examined in more detail and its effect to be realized on the circuit analyzed.

Basic principles
The ideal standard operational amplifier is an amplifier with two inputs and (normally) one output (Figure 14). The inputs are the non-inverting input and the inverting input. The differential voltage U_D is amplified and then made available at the output as the output voltage U_A. The following formula applies:

$$U_A = A_D\, U_D.$$

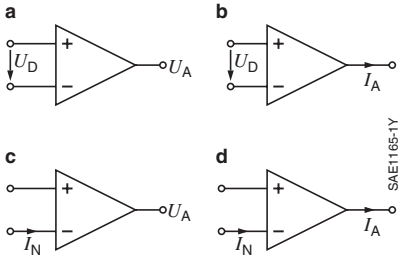
A_D is the open-loop gain. The operational amplifier is connected to a positive supply voltage and to a negative supply voltage with regard to the frame potential. In the case of a unipolar supply, the negative supply voltage can be applied to the frame potential. The supply voltages are not usually indicated in many circuit diagrams. They are however very much needed in order to guarantee the energy supply to the operational amplifier.

Figure 14: Basic diagram of an operational amplifier
+ Non-inverting amplifier input,
– Inverting amplifier input,
U_D Differential voltage between the two input potentials U_P and U_N where $U_D = U_P - U_N$,
U_A Output voltage,
U_{CC} positive supply voltage,
U_E negative supply voltage.



The following variants of the operational amplifier are common (Figure 15):
– "Normal" operational amplifier (VV operational amplifier) with voltage inputs and voltage output.
– Transconductance amplifier (VC operational amplifier) with voltage inputs and current output.
– Transimpedance amplifier (CV operational amplifier) with current inputs and voltage output.
– Current amplifier (CC operational amplifier) with current inputs and current output.

Figure 15: Operational-amplifier types (diagram)
a) Normal operational amplifier (VV) with voltage input and voltage output.
b) Transconductance amplifier (VC) with voltage input and current output.
c) Transimpedance amplifier (CV) with current input and voltage output.
d) Current amplifier (CC) with current input and current output.



As a rule, the VV operational amplifier is used; this will now be explained in more detail in the following. Because the circuit arrangement is of crucial importance to the function of an operational amplifier, this will be addressed in more detail first. The distinction between positive feedback and negative feedback is important here. Furthermore, an ideal operational amplifier will be assumed when deriving the correlations.

Circuit arrangement: negative and positive feedback

Negative feedback counteracts the cause. In the operational amplifier, a connection is required from the output to the inverting input for this purpose (Figure 16). This connection can be implemented by a network. The cause of a change in the output voltage U_A is always a change in the differential voltage U_D at the input; negative feedback therefore always acts in such a way that the voltage U_D becomes small and ideally zero.

Positive feedback, in contrast to negative feedback, supports the cause of the change at the output. U_A is thus amplified by positive feedback, i.e. U_D increases as U_A changes and is therefore always not equal to zero. In this way, the output voltage U_A can assume only two stationary values, namely the maximum value or the minimum value.

From a control-engineering standpoint, a negative feedback results from the operational amplifier and the feedback as shown in Figure 17. Taking a high gain A_D into consideration, it follows:

Figure 16: Negative and positive feedback
U_D Differential voltage,
U_A Output voltage.

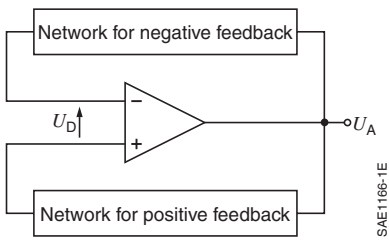
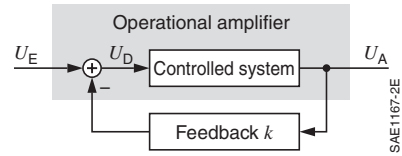


Figure 17: Negative feedback
U_E Input voltage as setpoint value,
U_A Output voltage,
U_D Differential voltage is obtained at the summing point $\oplus$ $U_D = U_E - k\,U_A$.
Controlled system with gain A_D,
Feedback with gain k.



$U_A = A_D U_D = A_D (U_E - k U_A)$

and for the total gain

$$A = \frac{U_A}{U_E} = \frac{A_D}{1 + k A_D} \approx \frac{1}{k} .$$

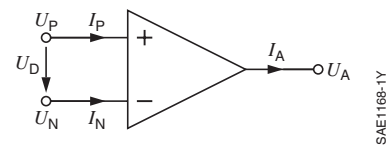
It thus becomes clear that, in spite of a very high open-loop gain A_D on the part of the operational amplifier, with the aid of negative feedback a finite gain A can be set with the negative-feedback network. This is explained below in greater detail using examples.

Ideal and real operational amplifier
First, the properties of an ideal operational amplifier as shown in Figure 18 will be summarized. For further information, see [1].
- Common-mode input resistance between, in each case, one input and ground, where: $r_{GL_P} = U_P/I_P$; $r_{GL_N} = U_N/I_N$. Generally speaking, r_{GL} can be ignored.
- Differential input resistance between the two inputs; here: $r_D = (U_P - U_N)/I_P$. r_D is increased by negative feedback.
- Output resistance, differential quantity $r_A = dU_A/dI_A$. r_A is decreased by negative feedback.
- Offset voltage U_{OS}: Characteristic quantity for describing the fact that even in the event of a short circuit between the two inputs (i.e. $U_D = 0$) the output voltage U_A is not equal to zero.
- Common-mode rejection ratio (CMRR): This quantity describes the change in the output voltage U_A if the two input voltages U_P and U_N change simultaneously (in the case of periodic input signals co-phasally), i.e. U_D remains constant.
- Power-supply rejection ratio (PSRR): Change in the output voltage U_A on account of a change in the supply voltages.

The essential idealizations are:
- The open-loop gain A_D approaches infinity; in the case of negative feedback the following applies: $U_D = 0$.
- The input currents I_N and I_P each approach zero.
- If I_N and I_P each approach zero, it follows that the common-mode and differential input resistances approach infinity.
- The offset voltage U_{OS} approaches zero.
- The output resistance R_A approaches zero.
- The common-mode rejection ratio (CMRR) approaches infinity, i.e. in the case of an equal and co-phasal change in the voltages U_P and U_N, U_A remains unchanged.
- The power-supply rejection ratio (PSRR) approaches infinity, i.e. in the case of a change in the supply voltage, U_A does not change.
- The behavior is not dependent on the frequency.

In reality, the above-mentioned idealizations do not apply entirely:
- The open-loop gain A_D is in the range of 10^4 to 10^7.
- The input currents I_N and I_P are in the range of 10 pA to 2 µA.
- The common-mode input resistance is in the range of 10^6 to 10^{12} Ω, the differential input resistance at up to 10^{12} Ω.
- The output resistance R_A is in the range of 50 Ω to 2 kΩ.
- The common-mode rejection ratio (CMRR) is in the range of 60 to 140 dB.
- The power-supply rejection ratio (PSRR) is in the range of 60 to 100 dB.
- The behavior is dependent on the frequency (low-pass behavior).

Figure 18: Ideal operational amplifier
U_D Differential voltage between the two input potentials U_P and U_N where $U_D = U_P - U_N$,
I_P, I_N Input currents,
U_A Output voltage,
I_A Output current.



Basic circuits
The external circuit arrangement of an operational amplifier determines the behavior of the entire circuit. Here, negative feedback plays the dominant role, since it enables the gain to be set exactly through the choice of resistances. The function shall now be explained by reference to several examples.

Inverting amplifier
The basic circuit for an inverting amplifier is shown in Figure 19. The name can be attributed to the negative gain, i.e. in the case of a periodic input voltage, the output voltage U_A is always in phase opposition to the input voltage U_1. In the following, it is important that, on account of the negative feedback and the high open-loop gain A_D, the differential voltage U_D at the input be constantly zero, since the non-inverting and inverting inputs are kept at the same potential. Since, on account of the negative feedback, the differential voltage U_D is regulated to zero, this is referred to as a "virtual short circuit". "Virtual ground" is also referred to here, as the inverting input is actively kept at zero (i.e. at frame potential). In addition, the input currents are ignored, in particular $I_N = 0$ is set. The following applies:

$$I_{R1} = \frac{U_1}{R_1} \text{ and } I_{R2} = -\frac{U_A}{R_2}.$$

Where $I_{R1} = I_{R2}$ it then follows:

$$U_A = -\frac{R_2}{R_1} U_1 .$$

The output voltage U_A is thus directly dependent on the input voltage U_1 and on the choice of resistances R_2 and R_1.

Non-inverting amplifier
The non-inverting amplifier can be handled along similar lines to the inverting amplifier (Figure 20). Due to negative feedback, $U_D = 0$. Where $I_{R1} = I_{R2}$, it is possible in accordance with the voltage divider – consisting of R_1 and R_2 – to calculate the voltage.

$$U_1 = \frac{R_1}{R_1 + R_2} U_A .$$

This means that

$$U_A = \frac{R_1 + R_2}{R_1} U_1 = \left(1 + \frac{R_2}{R_1}\right) U_1 .$$

The output voltage U_A is here also directly dependent on the input voltage U_1 and the choice of resistances R_2 and R_1; however, here the gain U_A/U_1 has at least the value one; U_A and U_1 are in phase.

Figure 19: Inverting amplifier
U_1 Input voltage,
U_D Differential voltage,
U_A Output voltage,
R_1, R_2 Network resistances,
I_{R1}, I_{R2} Currents in R_1 and R_2,
I_N Input current.

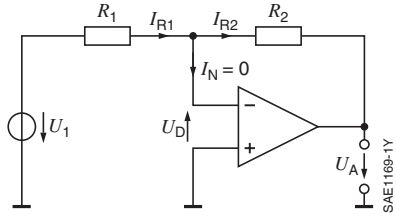
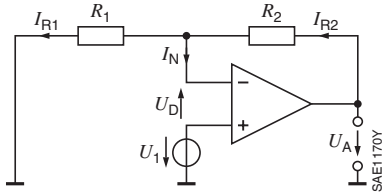


Figure 20: Non-inverting amplifier
U_1 Input voltage,
U_D Differential voltage,
U_A Output voltage,
R_1, R_2 Network resistances,
I_{R1}, I_{R2} Currents in R_1 and R_2,
I_N Input current.



A special case of the non-inverting amplifier is the isolation amplifier or impedance transformer. If R_1 assumes an infinitely high value (open loop) and R_2 is set equal to zero (short circuit) (Figure 21), then the gain is equal to one (i.e. $U_A = U_1$).

The advantage of this circuit is the property to the effect that the input-voltage source U_1 is not loaded with the internal resistance R_E, as the input current I_P is approximately equal to zero. This results in a negligible voltage drop over R_E and because $U_D = 0$, the input voltage U_1 is available at the output of the operational amplifier as U_A. This is important particularly for conditioning sensor signals, as here the sensor output voltage may not be loaded in many cases, i.e. any current flow from the sensor element can reduce the tappable voltage significantly.

Subtracting amplifier

The subtracting amplifier (Figure 22) can be considered as a common variant of the two previously mentioned circuits. On account of the superposition theorem (superposition principle), the relationship between the output voltage U_A and the input voltages U_1 and U_2 can be derived.

$$U_A = \frac{R_2}{R_1}(U_2 - U_1) .$$

Instrument amplifier

Particularly in sensor systems, differential voltages must often be picked off at bridge circuits and amplified without the sensor voltage or the bridge voltage being subjected to an unacceptably high load. This can be implemented by means of a high-resistance voltage tap. An instrument amplifier, which outputs the difference between two potentials U_2 and U_1 as amplified output voltage U_A, can be used for this purpose. The instrument amplifier can be subdivided into two parts: the pre-amplifier and a subtracting amplifier (Figure 22) with further amplification. Figure 23 shows the basic input circuit for the pre-amplification of an instrument amplifier.

According to the rule of negative feedback, the potential difference between the inverting and non-inverting inputs is equal to zero. In each case, the current I can flow through the resistors R and R', as the input currents I_{N1} and I_{N2} can be ignored. The following applies:

$$I = \frac{U_1 - U_2}{R'} = \frac{U_{A1} - U_{A2}}{2R + R'}, \text{ thus}$$

$$U_{A1} - U_{A2} = (U_1 - U_2)\left(\frac{2R}{R'} + 1\right).$$

Figure 21: Impedance transformer or isolation amplifier

U_1 Input voltage,
U_D Differential voltage,
U_A Output voltage,
R_E Input resistance,
I_P Input current.

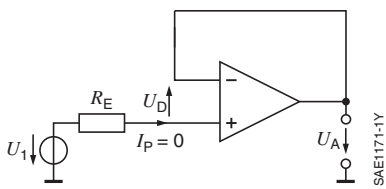
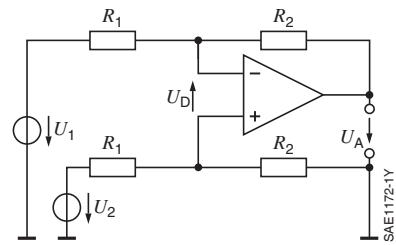


Figure 22: Subtracting amplifier

U_1, U_2 Input voltages,
U_D Differential voltage,
U_A Output voltage,
R_1, R_2 Network resistances.



In this way, the amplified difference between the two voltages U_1 and U_2 is obtained as the potential difference U_D between the two outputs of the two operational amplifiers. In order to output this voltage U_D as an output voltage referred to ground U_A, a subtracting amplifier can be connected in series (Figure 22), where U_{A1} instead of U_1 and U_{A2} instead of U_2 are supplied.

Important characteristic data
For many applications, the operational amplifiers must demonstrate properties which partly contradict each other. There are a large number of operational amplifiers, which are optimized for different fields of application. Basically, the data for specific operating points or operating ranges are specified.

Temperature range
In the field of consumer electronics, a temperature range of between 0 °C and 70 °C is customary. For the extended industrial field, a temperature range of between –20 °C and +70 °C is often specified; this range is required above all for devices which are used outside buildings. For military applications, a temperature range of –55 °C to +125 °C is specified. These demands do not, however, cover all the requirements for use in motor vehicles; for example, even higher temperatures occur in the engine compartment or in brake systems.

Offset voltage
The characteristic quantity for describing the fact that, even in the event of a short circuit between the two inputs (i.e. for $U_D = 0$), the output voltage U_A is not equal to zero is termed the offset voltage U_{OS}. This thus acts like a voltage applied from an external source U_D and is added to this. The offset voltage U_{OS} can, for example, be determined by the fact that at the input that voltage which sets the output voltage U_A equal to zero is determined (Figure 24). The offset voltage U_{OS} stems, among other things, from asymmetries in the internal circuit arrangement of the two inputs and is typically in the range of a few µV to a few mV.

However, as well as the value of the offset voltage U_{OS}, the temperature effect and long-term stability are also highly important. Some operational amplifiers offer the possibility of compensating the offset voltage by means of an external circuit arrangement – provided this is not already effected by internal circuit-engineering measures. It is important in this connection that the input voltage can also drift on account of the temperature effect; thus soldered junctions can represent thermocouples with a voltage of a magnitude of 10 to 100 µV/K.

Figure 23: Pre-amplification of an instrument amplifier
U_1, U_2 Input voltages,
I_{N1}, I_{N2} Input currents, I Current
R, R' Resistances
U_{A1}, $UA2$ Output voltages, in relation to ground.

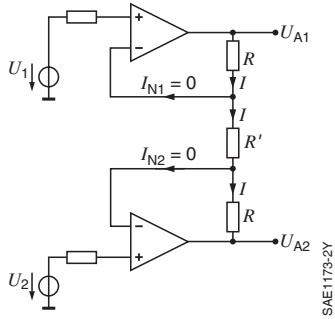
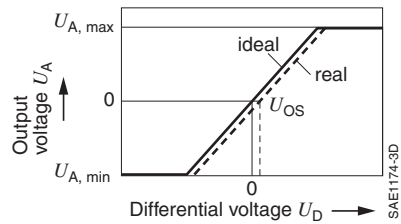


Figure 24: Offset voltage
U_A Output voltage,
U_D Differential voltage between the inputs,
U_{OS} Offset voltage,
$U_{A, MAX}$ Maximum output voltage,
$U_{A, MIN}$ Minimum output voltage.



Input resistances and currents
On account of the, as a rule, very low input currents I_N and I_P, very high input resistances, which are sometimes in the high megaohm range, are accordingly obtained. A distinction is made here between the common-mode input resistance (resistance between an input in each case and ground) and the differential input resistance between the two inputs.

The inputs of customary operational amplifiers form transistors; either bipolar transistors in which the base in each case is activated or MOS field-effect transistors in which the gate is recharged. The small input currents are explained in this way. When bipolar transistors are used, these are base currents and are in the range of μA. When MOSFET are used, the corresponding gate currents are obtained which are required to recharge the associated gate capacity. The latter are proportional to the operating frequency and are usually in the range of pA.

The input bias current can cause an input-voltage error in high-resistance circuits. This can be compensated for by connecting identical impedances to the two inputs, as then in each case the same voltage drops and the differential voltage U_D remains unaffected. Like the offset voltage, the input current can also drift over the temperature and the time.

Output resistance
The output of the operational amplifier can be described by a series connection of an ideal voltage source and a resistance; the latter is then the output resistance R_A. This resistance limits the output current. In general, operational amplifiers can drive output currents of 20 mA; there are also types with an output current of up to 10 A.

Slew rate
The slew rate (SR) denotes the maximum possible change in the output voltage U_A per period of time, i.e. the maximum value for dU_A/dt. The values for the slew rate for conventional operational amplifiers are in the range of less than 1 V/μs to more than 1 V/ns.

Noise
Noise can be described by specifying the noise-voltage density or the noise-current density. Usually the noise-voltage density U_R' is given in nV/$\sqrt{\text{Hz}}$.

The effective value of the noise voltage U_R (this also applies to the noise current) is obtained from the respective key figure multiplied by the root of the bandwidth B being considered:

$$U_R = U_R' \sqrt{B} \ .$$

For an amplifier circuit, the total effective noise-voltage density is obtained as the root of the sum of the squares of the effective values.

$$U'_{R,tot} = \sqrt{(U_{R,1})^2 + \text{to } (U_{R,m})^2} \ .$$

m here denotes the number of noise terms.

The noise is determined predominantly at the input of the operational amplifier. If JFET or MOSFET are used, low current but comparatively high voltage noise is obtained. The behavior is reversed in operational amplifiers which are based on bipolar transistors (see [1] and [4]).

Push-pull drivers

Operational amplifiers usually exhibit a maximum output current of approximately 20 mA. If a larger current is needed for example to activate power components or actuators (e.g. DC electric motors), this can be achieved with a driver stage – stages of this kind are also know as push-pull circuits, push-pull stages, or push-pull drivers. It is important in this connection that in some cases a large current I_{load} flows through the load (ohmic resistance R_{load}, Figure 25) in both directions. This is made possible by the use of two power transistors (in Figure 25 T_1 as *npn* and T_2 as *pnp* transistor); MOSFETs can also be used instead of bipolar transistors.

Because of the negative feedback the output voltage U_{OUT} is always equal to U_{IN} (Figure 26) and consequently for an ohmic load the output current I_{load} follows the input voltage U_{IN}.

The two transistors T_1 and T_2 in Figure 25 are activated in such a way that they never conduct simultaneously (push-pull operation) because this could lead to a short-circuit of the supply voltages $+U_V$ and $-U_V$. Since when the current I_{LOAD} is reversed through the load resistance R_{LOAD} first one transistor and then immediately afterwards the other transistor must conduct, this results in the voltage U_{OPV_out} in the circuit according to Figure 25 jumping abruptly from for example –0.7 V to +0.7 V, as shown in Figure 26. This can give rise to problems in applications with high dynamic response.

Push-pull circuits are frequently also used in driver circuits for IGBTs (insulated-gate bipolar transistors); a positive voltage (frequently +15 V) and a negative voltage (frequently –15 V) are made available here for the activation of IGBTs. The driver output stage is as a rule supplied with a metallically separated voltage via an insulated DC/DC converter. In addition, the driver circuit takes on various protective functions such as overcurrent detection or even interlocking logic.

Figure 25: Push-pull stage
Fundamental structure with two bipolar transistors.
T_1 *npn* bipolar transistor,
T_2 *pnp* bipolar transistor,
OVP_1 Operational amplifier,
R_{IN} Input resistance,
R_{LOAD} Load resistance.
U_{IN} Input voltage,
U_{OUT} Output voltage,
I_{LOAD} Load current,
$+U_V$ Voltage supply, positive potential,
$-U_V$ Voltage supply, negative potential.

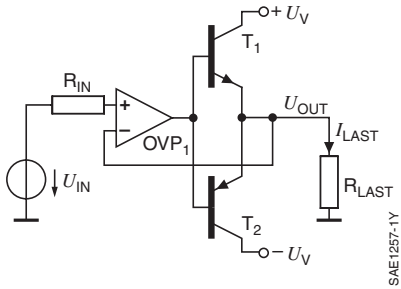
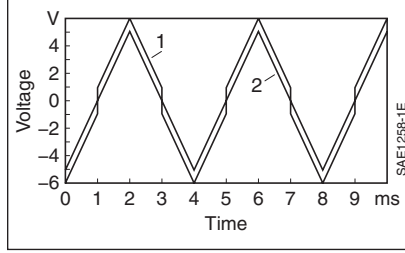


Figure 26: Characteristic curves of a push-pull stage
1 Voltage at output of OVP_1,
2 Output voltage ($U_{OUT} = U_{IN}$).



Monolithic integrated circuits

Monolithic integration
In monolithic integrated circuits (IC), components are assembled on a single piece of monocrystalline silicon (substrate). Semiconductor processes are used for creating layers (e.g. epitaxy), removing layers and changing material properties (e.g. doping). This technology enables complex circuits to be housed in the smallest of spaces.

Planar technology is based on the oxidation of silicon wafers, which is a relatively simple process, and the speed at which the dopants penetrate into silicon, which is exponentially greater than the speed at which they enter the oxide. Doping only occurs at locations where openings are present in the oxide layer. The specific design requirements of an individual integrated circuit determine the precise geometric configuration, which is applied to the wafer in a photolithographic process. All processing procedures (oxidizing, etching, doping, and depositing) progress consecutively from the surface plane (planar).

Planar technology makes it possible to manufacture all circuit components (e.g. resistors, capacitors, diodes, transistors) and the associated conductor strips on a single silicon chip in a unified manufacturing process. Monolithic integrated circuits are built from semiconductor components.

This integration generally comprises a subsystem within the electronic circuit and increasingly comprises the entire system (a system on a chip).

Because of the ever-increasing component density (integration density), the third dimension, i.e. the plane vertical to the surface, is also increasingly being utilized in the design. In this way, particularly in the field of power electronics, advantages such as lower resistances, smaller losses and thus also higher current densities can be achieved.

Integration level
The integration level is a measure of the number of function elements per chip. The following technologies relate to the level of integration (and chip surface):

- SSI (Small Scale Integration) with up to several hundred function elements per chip and a mean chip surface area of 1 mm^2. But the chip surface area can be very much larger in circuits with high power outputs (e.g. smart power transistors).
- MSI (Medium Scale Integration) with a several hundred to 10,000 function elements per chip and a mean chip surface area of 8 mm^2.
- LSI (Large Scale Integration) with up to 100,000 function elements per chip and a mean chip surface area of 20 mm^2.
- VLSI (Very Large Scale Integration) with up to 1 million function elements per chip and a mean chip surface area of 30 mm^2.
- ULSI (Ultra Large Scale Integration) with over 1 million function elements per chip (flash memory today contains up to 20 billion transistors per chip), a surface area of up to 300 mm^2 and the smallest structure sizes of less than 30 nm.

Computer-aided simulation and design methods (CAE and CAD) are essential elements in the manufacture of integrated circuits. Entire function modules are used in VLSI and ULSI, otherwise the time expenditure and failure risk would make development impossible. In addition, simulation programs are used to detect any defects made.

References
[1] U. Tietze, Ch. Schenk, E. Gamm: Halbleiter-Schaltungtechnik. 15th Edition, Verlag Springer Vieweg, 2016.
[2] A. Führer, K. Heidemann, W. Nerreter: Grundgebiete der Elektrotechnik, Volumes 1–2. 9th Ed., Carl Hanser Verlag, 2011.
[3] R. Ose: Elektrotechnik für Ingenieure. 5th Edition, Carl Hanser Verlag, 2013.
[4] R. Müller: Rauschen. 2nd Ed., Springer-Verlag, 2013.

Basic principles of electrical machines

Energy conversion

Electric machines are electro-magneto-mechanical energy converters in motor and generator operation. The supplied electrical energy W_{el}

$$W_{el} = W_m + W_v$$

is converted into magnetic energy W_m and lost energy W_v where

$$W_m = \int_0^{t_0} u_l i_l dt \quad \text{and} \quad W_v = \int_0^{t_0} i_l^2 R\, dt.$$

Where $u_l = d\Psi/dt$ the magnetic energy follows by means of integration

$$W_m = \int_0^{\Psi_0} i_l(\delta, \Psi)\, d\Psi.$$

Here δ is the air gap in the magnetic circuit, Ψ the magnetic flux interlinked with all the turns of a winding, and Ψ_0 the maximum occurring interlinked magnetic flux. In Figure 1 for example the energy W_m is accumulated in the space $V = A_z \delta$ above the tooth surface permeated by the flux Φ. Through discharging of the magnetic energy to a space coordinate Ω the magnetic force F_m follows

$$F_m = \frac{\partial W}{\partial \Omega}$$

in the direction of the space coordinate as a force on the boundary layer for the generation of motion. If, for example, the magnetic energy is discharged to a circumferential coordinate, this results in a magnetic force in the circumferential direction which through multiplication by radius r causes the torque $M = F_m r$. If the magnetomechanical energy conversion is effected through discharging of the magnetic energy to an angular coordinate α, torque M follows:

$$M = \frac{\partial W}{\partial \alpha}.$$

The section of a permanent-magnet motor is shown in Figure 1. The stator consists of the magnetic yoke and the permanent magnet. The rotor is formed by the rotor-tooth surface A_z, the rotor tooth, and the winding, which is permeated by the magnetic flux and causes a torque-producing reluctance force (tangential force F_t) at the tooth surface. The normal force F_n acting on the tooth surface counterbalances together with the normal force of the opposite tooth on the shaft.

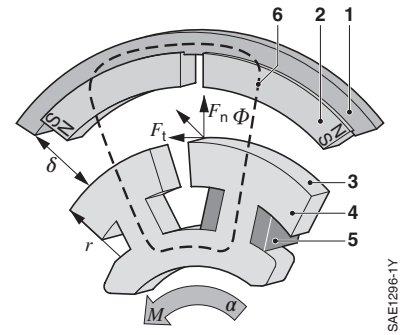
With regard to forces, a distinction is made between the following:
– magnetic force which acts on a boundary layer consisting of air and iron,
– Lorentz force on moving charge carriers in the magnetic field.

Application examples for the first case are commutator machines and switched reluctance motors. Application examples for the second case are commutator machines with ironless rotors and commutator disk-rotor machines.

For further references, see [1] and [2].

Figure 1: Sectional view of an internal-rotor motor

1 Magnetic yoke,
2 Permanent magnet,
3 Rotor-tooth surface A_Z,
4 Rotor tooth,
5 Armature winding,
6 Magnetic flux Φ.
r Rotor radius,
F_t Tangential force (reluctance force),
F_n Normal force, M Torque,
α Angle of rotation,
δ Air gap in the magnetic circuit.



Systematics of rotating electric machines

The energization of the windings of electric machines determines their properties. A distinction is made between self-commutated (position-commutated, field-commutated) and externally commutated (line-commutated, frequency-commutated) motors.

The first group includes DC motors in which cyclic activation of the winding phases in the rotor is effected on a self-regulated basis via the mechanical (or electronic) commutator. Magnetic excitation is effected either with an electromagnet (DC and AC operation of the motor possible) or by means of permanent magnets (only DC operation of the motor possible).

In externally commutated motors (e.g. asynchronous motor) the phases are switched by the supply system or the control electronics. Table 1 shows a systematization of electric motors with regard to their commutation performance. DIN 42027 [3] sets out motor systematics with regard to power supply.

Definitions of geometric quantities

Figure 2 depicts a 4-pole (number of pole pairs $p = 2$) machine cross-section. The quantities are named in part by DIN EN 60027-4 [4]. This standard gives names and letter symbols for quantities and units.

A pole pair consists in each case of a north pole and a south pole. The magnetic flux that is set emerges from the north pole through the rotor via the air gap into the south pole.

Figure 2: Stator section and rotor section with dimensioning

1 Stator,
2 Rotor (with radius r_R).
τ_p Pole pitch (= $\pi d/2p$),
p Number of pole pairs,
τ_Q Slot pitch (= $\pi d/Q$),
Q Number of slots,
b_L Pole-gap width,
b_P Pole-shoe width,
r_S Stator radius,
δ Air gap,
Φ Magnetic flux.

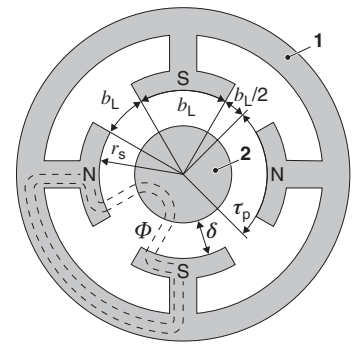


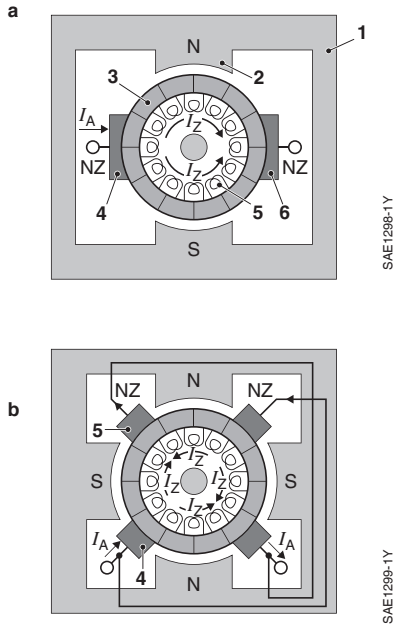
Table 1: Systematics of electric motors

Self-commutated machines			Externally commutated motors	
Mechanical commutator		Electronic commutator	Load-dependent rotational speed	Frequency-dependent rotational speed
AC motors	DC motors	Electronically commutated motors (EC motors)	Asynchronous motors	Synchronous motors
Commutator motors, universal motors	Series-wound, shunt-wound motors	Block-, sinusoidally commutated motors	Rotating-field motors, cage motors	Rotating-field motors

Direct-current machines

In direct-current machines motor operation is frequently preferred to generator operation. These are used for example as drives for electric fuel pumps, fans, starters, windshield wipers, and power-window units. Figure 3a shows a two-pole motor and Figure 3b a four-pole motor. In higher-pole motors a number of shorter magnetic circuits are generated which facilitate greater utilization of the magnet volume. The motors consist of the magnetic yoke, the poles, copper segment, inlet brush, winding phase, and outlet brush, through which the armature current I_A passes which in turn is split into the branch-circuit currents I_Z.

Figure 3: DC machines
a) Two-pole motor (1 pole pair),
b) Four-pole motor (2 pole pairs).
1 Magnetic yoke, 2 Pole,
3 Copper segment, 4 Inlet brush,
5 Winding phase, 6 Outlet brush.
I_A Armature current, I_Z Branch-circuit current.
NZ Neutral zone.



Direct-current machines with excitation by electromagnets are classified according to their winding behavior. In a series-wound motor the armature and excitation windings are connected in series. In a shunt-wound motor the armature and excitation windings are connected in parallel. The terminal markings of direct-current machines comply with DIN EN 60034 Part 8 [5] and DIN EN 60617 Part 6 [6].

Commutator voltage

Figure 3 is simplified down to the equivalent motor circuit diagram Figure 4, consisting of commutator and two phases reduced to one turn. The equivalent electric circuit diagram Figure 5 is derived from this. Figures 5a and 5b show motor operation, Figures 5c and 5d show generator operation. The voltage relationships that are obtained can be taken from Table 2.

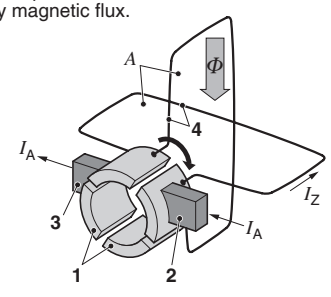
The induced voltage U_L is according to Faraday's law

$$U_L = \frac{d\Psi(I,\delta)}{dt}. \qquad (1)$$

With the total differential (differentiation of a function according to all its independent variables) of the interlinked, current- and air-gap-dependent magnetic flux Ψ

$$d\Psi = \frac{\partial\Psi(I,\delta)}{\partial I}\, dI + \frac{\partial\Psi(I,\delta)}{\partial\delta}\, d\delta \qquad (2)$$

Figure 4: Equivalent motor circuit diagram
1 Copper segment, 2 Inlet brush,
3 Outlet brush, 4 Winding phase.
I_A Armature current, I_Z Branch-circuit current.
Φ Magnetic flux,
A Area permeated by magnetic flux.



the induced voltage follows again

$$U_L = \frac{\partial \Psi(I,\delta)}{\partial I}\,\frac{dI}{dt} + \frac{\partial \Psi(I,\delta)}{\partial \delta}\,\frac{d\delta}{dt}, \qquad (3)$$

where the first term becomes zero due to DC energization. This leaves the second term of Faraday's law, which is further developed with the number of turns N, the flux Φ, the flux density B, and the area A:

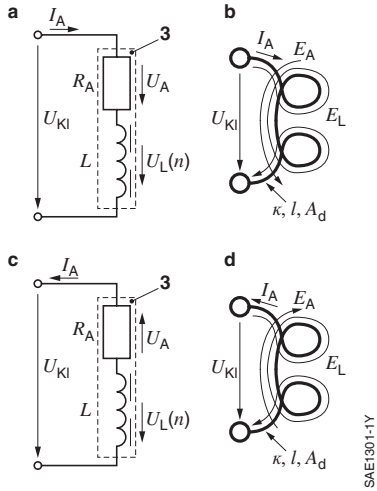
$$U_L = N\frac{\partial \Phi}{\partial \delta}\,\frac{d\delta}{dt} = NB\frac{dA}{dt}. \qquad (4)$$

Table 2: Voltage relationships

Motor operation	Generator operation
$U_{Kl} = U_A + U_L$	$U_{Kl} = U_A - U_L$
$U_{Kl} = E_A l + E_L l$	$U_{Kl} = E_A l - E_L l$
$U_{Kl} = (E_A + E_L)\, l$	$U_{Kl} = (E_A - E_L)\, l$

Figure 5: Equivalent electric circuit diagram for motor and generator operation

I_A Armature current,
U_{Kl} Terminal voltage,
R_A Armature resistance,
L Armature inductance,
U_A Voltage drop via armature resistance R_A,
U_L Induced voltage,
E_A Electric field strength along the winding wire,
E_L Field strength induced in the winding,
l Wire length,
A_d Wire cross-sectional area,
κ Specific electric conductivity.



With the replacement of the differential function by a differential function using the period T and taking into account that for the arrangement in Figure 4 with $k = 4$ segments for commutation only $T/4 = T/k$ time remains, the induced voltage follows again

$$U_L = NB\,\frac{A}{\frac{T}{k}} = NBAfk \qquad (5)$$

as a function of the rotational frequency f [s^{-1}]. An extension with 60 and conversion with $c_1 = k \cdot 60$ leads to

$$U_L = \Psi n k \cdot 60 = L I n c_1 = \Psi n c_1 \,(n \text{ in } \text{min}^{-1})$$

and thus to the rotational-speed-proportional induced voltage with which the mesh equation of Figure 5

$$U_{Kl} = R_A I_A \pm \Psi n c_1 \qquad (6)$$

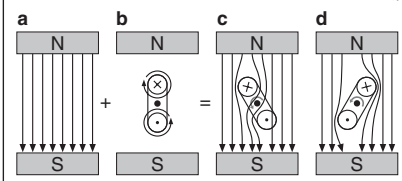
is fulfilled. Here "+" applies to motor operation and "–" to generator operation. n specifies the revolutions per minute.

Commutation

Only the commutator facilitates rotation of the armature and serves to reverse the direction of the current in a winding phase ("commutate" means to change quantities, to transpose with each other, to change the direction of electric current). Commutation is therefore the reversal of the current direction.

Figure 6: Superposition of fields during commutation

a) Undisturbed stator field distribution,
b) Undisturbed armature field distribution,
c) Superposition of field distribution from a and b,
d) Reversal of current direction while maintaining the direction of armature rotation.



The undisturbed stator field distribution is shown in Figure 6a. The undisturbed armature field distribution is drawn in Figure 6b. The two cases are superposed in Figure 6c. This generates a resulting force which allows the armature to rotate to the left. The current direction is reversed by a commutator in Figure 6d while maintaining the direction of armature rotation.

The rotor from Figure 3a is shown enlarged in Figure 7. The armature current I_A is split in each case into branch-circuit currents I_Z. The branch-circuit currents flows into the winding phases of the north and south poles. Beneath the pole faces the current direction in the winding phases is therefore always the same. The current direction changes in each case only in the winding phase to be commutated. The commutation of a winding phase at the inlet brush is given as an example of a commutation. The coils to be commutated are in the neutral zone (NZ). With an accepted right-hand rotation of the rotor the winding phase runs underneath the inlet brush (Figure 7, position 1). Its present current direction is marked with an arrow. In due course the winding phase is short-ed-circuited by the inlet brush (Figure 7, position 2). In position 3 the energization direction of the branch-circuit current was reversed (commutation), which is marked by an arrow.

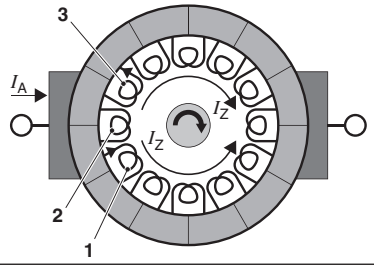
The time process of commutation within the commutation period T_C can be seen in Figure 8. The start of commutation is marked with t_{CB} and the end of commutation with t_{CE}. The ideal case is complete commutation following the shape of curve 1. Here the effect of the inductive voltages is canceled out by the commutating pole (see following section). The voltage induced by the commutating field is equal to the reactance voltage. Curve shape 2 is called under-commutation. The coil current I decreases from $+I_Z$ only slowly and reaches the current level $-I_Z$ only shortly before the time t_{CE}. The cause of this is an excessively underdimensioned induced commutating-pole voltage.

The opposite to this is curve shape 3, which describes an over-commutation. This is achieved by an excessively high induced commutating-pole voltage.

Basically, the abrupt current change loads the contact system consisting of brush and segment.

The designations have been taken from DIN 1304, Part 7 [7].

Figure 7: Commutation process in the rotor
1 Winding phase under the inlet brush,
2 Winding phase short-circuited by inlet brush,
3 Reversal of energization.
I_A Armature current.



Commutating and compensation winding

Compensated machine

Figure 9 depicts a machine provided with a commutating and compensation winding for higher power output (approximately > 5 kW).

Figure 8: Commutation
1 Ideal case of complete commutation,
2 Under-commutation,
3 Over-commutation.
I_Z Branch-circuit current,
T_C Commutation period,
t_{CB} Start of commutation,
t_{CE} End of commutation.

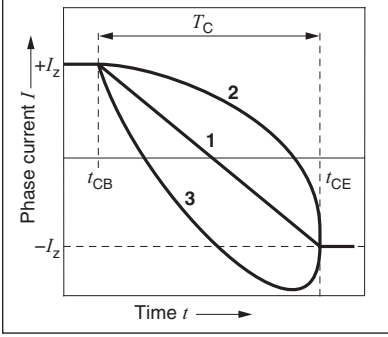
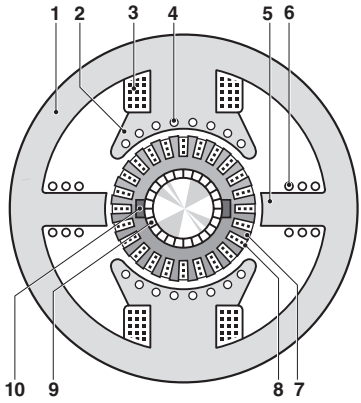


Figure 9: Structure of the two-pole DC machine (cross-section)

1 Stator, 2 Exciter pole (pole shoe),
3 Exciter winding,
4 Compensation winding (preferably on machines with higher power output),
5 Commutating pole,
6 Commutating winding, 7 Rotor,
8 Rotor winding, 9 Commutator lamina,
10 Commutator brush (position in the neutral zone).



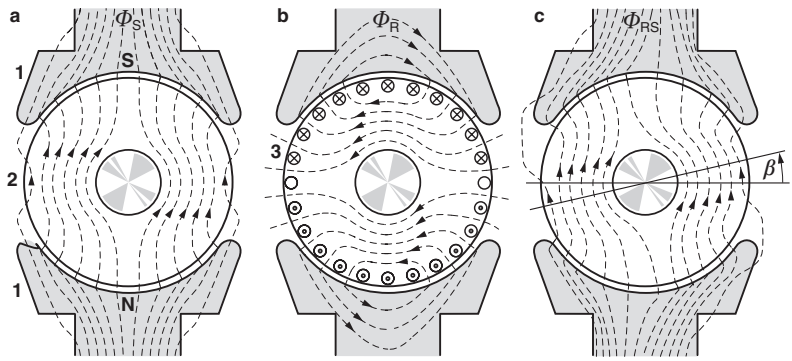
Field distribution

The main field that penetrates the rotor without hindrance when no current is present has a symmetrical distribution (Figure 10a). Likewise, there is a symmetrical flux division when current is applied to only the rotor (Figure 10b). If both fields are superpositioned, the neutral zone is deflected by the angle β (Figure 10c). This magnetically neutral zone thus no longer corresponds to the geometrically neutral zone (position of the commutator brushes). In the geometrically neutral zone, this creates a magnetic field through which a voltage is induced during the commutation process in the coil to be commutated, creating brush sparking (break sparks) between the brush and the running commutator lamina. To prevent this, another voltage is induced during commutation in the coil concerned where the amplitude and direction cancel out the effect of the originally induced voltage.

This is achieved by means of the commutating winding (Figure 9). The commutating winding is connected in series to the rotor winding. It uses rotor retroaction to counteract a shift in the magnetically neutral zone. In the case of motors with-

Figure 10: Superposition of fields

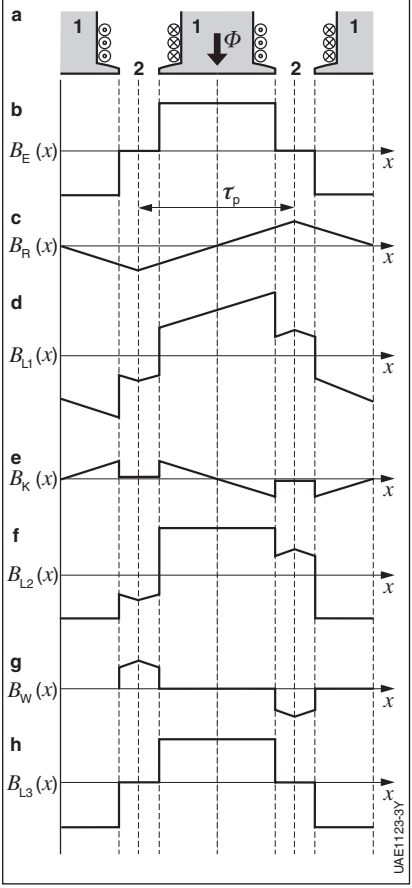
a) Main field: exciting current switched on, rotor current switched off,
b) Armature quadrature-axis field: exciting current switched off, rotor current switched on,
c) Complete field: superposition of main and armature quadrature-axis field; magnetic neutral zone is deflected by angle β.
1 Exciter pole (stator pole, main), 2 Rotor, 3 Rotor winding.
Φ_S Magnetic stator flux (magnetic yoke via the motor housing),
Φ_R Magnetic rotor flux (magnetic yoke via the exciter pole or motor housing),
Φ_{RS} Magnetic flux of the complete field.



out commutating windings, the brushes have to be shifted into the magnetically neutral zone.

The main field distortion that occurs in the area of the pole shoe leads to a reduction in the available pole surface, combined with an increase in the magnetic resistance. This is why larger machines are given a compensation winding that is integrated in the pole shoes (Figure 9). The compensation winding is connected in series with the rotor winding and its dimensions are such that it compensates for the rotor quadrature-axis field.

Figure 11: Effect of the compensation and commutating winding
a) Pole arrangement,
b) Distribution of the excitation field $B_E(x)$,
c) Distribution of the rotor quadrature-axis field $B_R(x)$,
d) Superposition of $B_E(x)$ and $B_R(x)$,
e) Compensation induction $B_K(x)$,
f) Superposition of $B_E(x)$, $B_R(x)$ and $B_K(x)$,
g) Commutating induction $B_W(x)$,
h) Superposition of all field distributions.
1 Pole shoe, 2 Neutral zone.
Φ Stator flux.



Effect of the commutating and compensation winding

The sequence of images in Figure 11 describes the effect of both windings. The field distributions in the air gap are shown. The pole arrangement with winding and the neutral zone can be seen in Figure 11a. The distribution of the excitation field $B_E(x)$ below the pole shoe as well as the pole pitch τ_P are shown in Figure 11b. Figure 11c shows the rotor quadrature-axis field distribution $B_R(x)$. The superposition of both field distributions can be seen in Figure 11d. The compensation induction $B_K(x)$ (Figure 11e) as well as the superposition that took place with Figure 11d are shown in Figure 11f. If the commutating induction $B_W(x)$ in Figure 11g is superposed with the field distribution of Figure 11f, this results in the desired field distribution according to Figure 11h.

Torque and power

The power input (+) or output (–) by the motor at the terminal is

$$P_{el} = P_{th} \pm P_{em} = I_A^2 R_A \pm U_L I_A, \qquad (7)$$

where P_{em} corresponds to the internal power and P_{th} to the thermal power, R_A is the armature resistance, and I_A the armature current. The internal mechanical motor torque M corresponds to

$$M = \frac{P_{em}}{\omega} = \frac{c_1 \Psi n I_A}{2\pi f} = \frac{c_1}{2\pi 60} \Psi I_A. \qquad (8)$$

For operation at constant rotational speed and mechanical load torque M_{mech} the following applies

$$M = M_{mech}.$$

Shunt-wound machine

A feature of the shunt-wound machine is that the rotor winding is connected parallel to the exciter winding. The following terminal markings (Figure 12) as set out in DIN EN 60034-8 [5] apply to the shunt-wound machine:

– A indicates the armature winding,
– B the commutating winding,
– C the compensation winding, and
– E the shunt field winding.

By transposing equation (6) in accordance with the rotational speed n the speed/current equation follows

$$n(I_A) = -\frac{R_A}{c_1 \Psi} I_A + \frac{U_{KI}}{c_1 \Psi} \qquad (9)$$

as a line equation with a negative slope and axis intercepts

$$n_0(I_A = 0) = \frac{U_{KI}}{c_1 \Psi};\ I_{An}(n = 0) = \frac{U_{KI}}{R_A}. \qquad (10)$$

Here n_0 is the idle speed and I_{An} the starting current. Through the conversion of equation (8) in accordance with the current and insertion in equation (9) the speed/torque equation follows as a line equation

$$n(M) = -\frac{R_A c_2}{{c_1}^2 \Psi^2} M + \frac{U_{KI}}{c_1 \Psi},\ c_2 = 2\pi \cdot 60 \qquad (11)$$

The axis intercepts are given with

$$M(n = n_0) = 0,\ M_{An}(n = 0) = \frac{U_{KI}}{R_A} \frac{c_1}{c_2} \Psi \qquad (12)$$

In Figure 13 the speed is shown against the armature current and the motor torque with four quadrants. Here, motor operation differs from generator operation in each case through the direction of rotation and torque. The speed points of the machine can follow by changing the terminal voltage U_{KI} according to equation (9), which brings about a parallel shift of the straight line in Figure 13. A change of the excitation current separate from the armature current permits according to equation (9) a change of the flux Ψ and thus also a speed change. This measure is called field weakening.

The straight-line slope of equation (9) is changed by a series connection of armature resistance R_A with a series resistance R_V in the armature circuit (Figure 14a). Speed curves with the series resistance as the parameter can be seen in Figure 14b.

Figure 12: Terminal markings of the shunt-wound machine

A Rotor winding, B Commutating winding, C Compensation winding, E Exciter winding, M Motor.

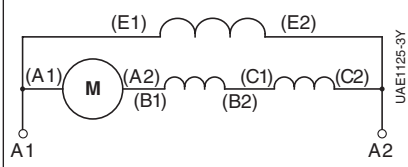


Figure 13: Four-quadrant operation of the shunt-wound machine

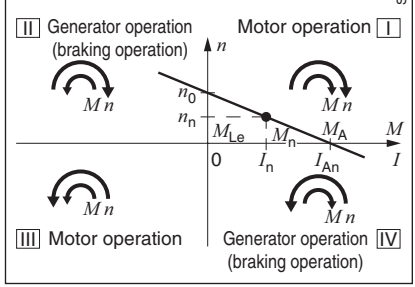
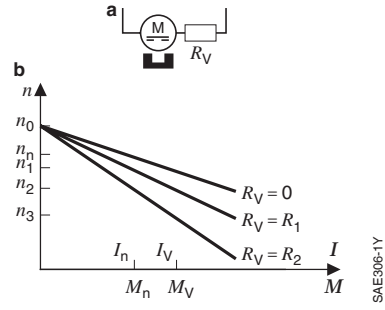


Figure 14: Speed points by means of armature series resistance

a) Series connection of armature resistance and series resistance,
b) Speed curves with the series resistance as parameter.



Series-wound machine

In the case of the series-wound machine, the commutating, compensation, exciter, and rotor winding are connected in series (Figure 15). D designates the exciter winding of the series-wound machine. To determine the operating characteristics, the ohmic resistances of the windings are added up to form the resistance R_A. The relationship $\Psi = L I$ and equation (8) inserted in equation (9), the speed/torque equation follows the series-wound machine:

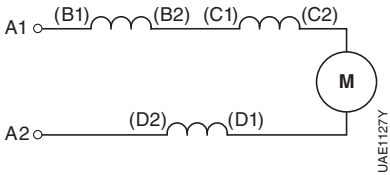
$$n(M) = \frac{1}{c_1 L}\left(U_{Kl}\sqrt{\frac{c_1 L}{c_2 M}} - R_A\right). \qquad (13)$$

Because of the torque in the denominator the motor achieves reaches high speeds at low load an may never be operated completely without base load. Owing to the absence of external excitation the series-wound commutator machine can only function as a motor. The speed/torque characteristics limits the use of the motor. Areas of application must therefore always provide for a base load (friction torque) as in a vacuum cleaner or in pumps. A small load change always brings about a large speed change. The speed-setting options of the series-wound machine are limited to changing the terminal voltage.

Figure 16 shows the speed/torque diagram of the series-wound machine with the terminal voltage U_{Kl} as the parameter.

Figure 15: Terminal markings of the series-wound machine

A Rotor winding, B Commutating winding, C Compensation winding, D Exciter winding, M Motor.



Asynchronous machine

The asynchronous machine is the main drive used in industry. In the automotive sector, for example, it is used in electrical power-assisted steering and in hybrid vehicles. The following section presents the operating concept of the asynchronous machine as an induction machine.

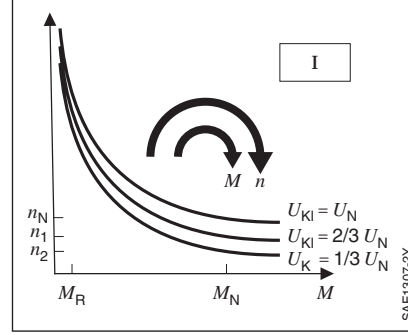
General setup

A distinction is made between external rotor and internal rotor machines. In the case of the external rotor machine, the rotor encloses the stator; on the internal rotor machine, the stator encloses the rotor.

The diagram of the principle (Figure 17) indicates the fundamental structure of an internal rotor asynchronous machine. The rotor consists in the simplest case of a short-circuited coil (short-circuit rotor). The stator consists of three coils with an iron core which are each assigned to a phase. The iron core consists of individual steel sheets insulated against one another to keep the eddy-current losses to a minimum. The rotating stator magnetic field induces in the short-circuited coil a current which for its part causes a magnetic field which is coupled to the rotating stator field and thus becomes effective in terms of torque.

Figure 16: Operation characteristic curves of the series-wound machine

n Motor speed, n_n Rated speed, U_{Kl} Terminal voltage, U_N Rated voltage, M_M Motor torque, M_R friction torque, M_N Rated torque.



Operating characteristics

The stator winding generates a rotating field with a three-phase alternating current. There is a speed difference between the rotating-field speed and the rotor speed that enables the induction of a magnetically effective current in the rotor, which in turn contributes to the generation of torque. The physical operating concept is based on the law of magnetic induction. Figure 18 shows the rotor as a simplified conductor loop on rotating mounts. The relative movement between the rotating stator field and the rotor is described by the slip angular frequency ω_S.

The magnetic field B_R rotating with the slip angular frequency induces according to Faraday's law in the cage rotor the voltage

$$\oint E\,ds = -\int\int \frac{dB}{dt}\,dA_S. \qquad (14)$$

E denotes the electric field strength along the short-circuit rods and bridges. There follows the development of the left term
– with material ($E = J/\kappa$),
– geometric ($\int E\,ds = E \cdot 2\,(l+2r)$, and
– electrical ($J = i_{ind}/A_{nom}$) quantities

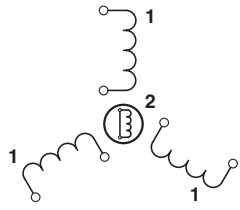
$$\oint E\,ds = \frac{2\,(l+2r)}{\kappa A_{nom}}\,i_{ind} = R_S\,i_{ind}, \qquad (15)$$

followed by the development of the right term using geometric ($A_s = 2\,l\,r$) and magnetic quantities in the frequency range

$$\int\int \frac{dB}{dt}\,dA_S = 2\,l\,r\,\hat{B}_R\,\omega_S \sin(\omega_S t) = \hat{u}_{ind}\sin(\omega_S t). \qquad (16)$$

Figure 17: Principle of a three-phase asynchronous motor

1 Fixed stator winding,
2 Rotor as rotatable short-circuited coil.
Axis of rotation perpendicular to the plane of projection.



Here, κ is the specific electric conductivity, J the current density, and R_S the conductor-loop resistance (short-circuit rods and bridges). By bringing together equations (15) and (16) and converting according to the current i_{ind} there follows

$$i_{ind} = \frac{\hat{u}_{ind}}{R_S}\sin(\omega_S t). \qquad (17)$$

With the conductor current the tangential force F_t (Lorentz force)

$$F_t = i_{ind}\,l\,\hat{B}_R \sin(\omega_S t) \qquad (18)$$

is obtained at the conductor loop. Together with the two rotor rods the torque M follows

$$M = 2\,F_t\,r = \frac{2\,\kappa A_{nom}\,\omega_S}{l+2r}\left(l\,r\,\hat{B}_R \sin(\omega_S t)\right)^2 \qquad (19)$$

with material, geometric, and magnetic quantities. Including the trigonometric function

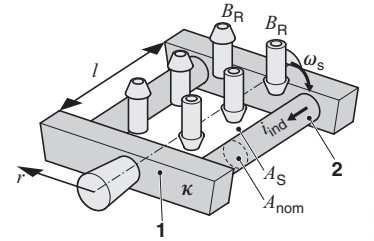
$$\sin^2(\omega_S t) = \frac{1}{2}\left(1 - \cos(2\,\omega_S t)\right)$$

the torque M follows again

$$M = \frac{\kappa A_{nom}\,\omega_S}{l+2r}\left(l\,r\,\hat{B}_R\right)^2 (1 - \cos(2\,\omega_S t)). \qquad (20)$$

Figure 18: Two-bar rotor as short-circuit rotor

1 Short-circuit bridge, 2 Short-circuit rod.
i_{ind} Induced current, l Short-circuit rod length,
r Rotor radius, A_{nom} Rod cross-sectional area,
B_R Magnetic field rotating with slip angular frequency,
A_S Cross-sectional area of conductor loop,
κ Specific electric conductivity of conductor loop,
ω_S Slip angular frequency (results from the relative movement between stator rotating field and rotor speed).



This is made up of a constant and a time-variable component of oscillating torque at twice the slip angular frequency. The amplitude of the oscillating torque corresponds to the amount of the constant component. The magnetic effect of the induced current in the conductor loop is taken into account as follows. The current causes, on account of Ampère's law, the induced magnetic field H_{ind} and the field strength the induced magnetic flux density B_{ind}:

$$H_{ind} = \frac{i_{ind} N}{l_H};\ B_{ind} = \mu H_{ind}. \quad (21)$$

Here, N is the number of conductors and l_H the field-line length. The power loss P_v obtained in the rotor is calculated at

$$P_V = R_S i_{ind}^2 = \frac{l_S}{\kappa A_{nom}} i_{ind}^2 \quad (22)$$

Figure 19 shows torque curves as a function of the slip angular frequency.

Synchronous machine

Synchronous machines are used preferably as claw pole generators. In motor operation, they are used for example in electrical power-assisted steering, to drive hybrid vehicles, and in electrically driven turbochargers.

General setup

In contrast to the asynchronous machine, in the synchronous machine the rotor rotates synchronously with the excitation field at the angular velocity $\omega_{\Phi S}$. The flux Φ_R generated by the rotor winding and the stator flux Φ_S superposition themselves in relation to the resulting flux Φ_{RS} (Figure 20):

$$\Phi_{RS} = \Phi_R + \Phi_S. \quad (23)$$

As the rotor and stator material are operated far below the magnetic saturation ($\mu_r \rightarrow \infty$), the air gap d between the rotor and stator as well as the angle α determine the magnetic circuit resistance R_m:

$$R_m = \frac{2\delta}{(\mu_0 A_R)} = \frac{2d}{(\mu_0 A_R \cos\alpha)}. \quad (24)$$

Figure 19: Torque curves of an asynchronous machine

1 Distribution under the influence of inductive reactance,
2 Distribution without the effect of inductive reactance.
ω_S Slip angular frequency,
ω_{Smax} Maximum possible slip angular frequency,
ω_K Breakdown angular frequency,
M Motor torque,
M_K Breakdown torque.

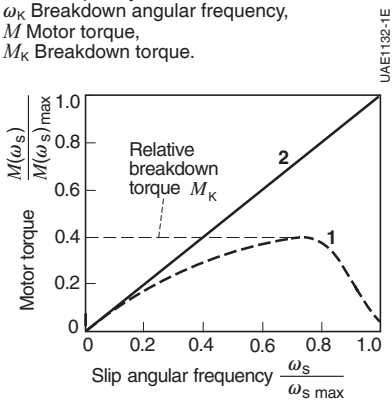
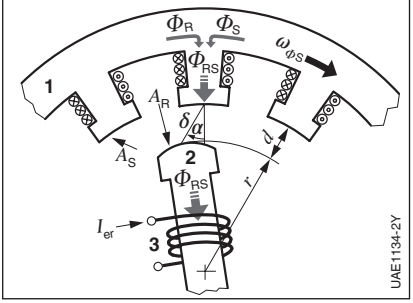


Figure 20: Basic design of the synchronous machine

1 Stator, 2 Rotor,
3 Rotor winding with N coils.
Φ_S Stator flux, Φ_R Rotor flux,
Φ_{RS} Superpositioned flux,
$\omega_{\Phi S}$ Angular velocity of the excitation field,
I_{er} Exciting current in the rotor,
A_S Magnetically effective stator surface,
A_R Magnetically effective rotor surface,
r Radius of the rotor,
d Distance between rotor and stator,
δ Air gap length, α Deflection angle.



The factor 2 results from the fact that there are two air gaps between the rotor and stator. If the motor provides torque, the rotor rotates with the angle α from its idle position (Figure 21).

The resulting flux Φ_{RS} is calculated as:

$$\Phi_{RS} = \frac{\Theta_{er}}{R_m} + \Phi_S . \quad (25)$$

With R_m from equation (24), this results in

$$\Phi_{RS} = \frac{\Theta_{er}\mu_0 A_R \cos\alpha + 2d\Phi_S}{2d} . \quad (26)$$

With $\Theta_{er} = NI_{er}$ the result is

$$\Phi_{RS} = \frac{NI_{er}\mu_0 A_R \cos\alpha + 2d\Phi_S}{2d} . \quad (27)$$

Θ_{er} is the magnetic rotor flooding and I_{er} is the exciting current fed to the rotor via the collector rings. The tangential force F_t affecting the torque is calculated using the Maxwell pole force formula

$$F_t = \frac{{\Phi_{RS}}^2}{\mu_0 A_R} \sin\alpha \quad (28)$$

Figure 21: Torque/load-angle characteristic curve
M_K Breakdown torque,
α_K Deflection on reaching the breakdown torque.

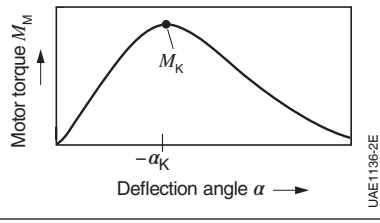
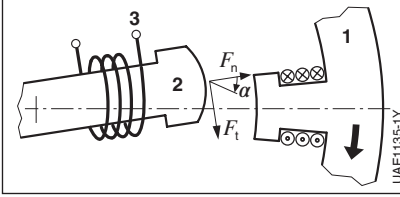


Figure 22: Forces at the rotor
1 Stator, 2 Rotor, 3 Rotor winding.
F_t Tangential force, F_n Normal force,
α Deflection angle.



[4, 5, 8]. The tangential force is used to calculate the motor torque M_M:

$$M_M = 2F_t r . \quad (29)$$

Equation (27) inserted in equation (28) and the result in equation (29) yields the following relationship:

$$M_M = \frac{r\sin\alpha}{2\mu_0 A_R d^2} \cdot [(NI_{er}\mu_0 A_R \cos\alpha)^2 + 4NI_{er}\mu_0 A_R d\Phi_S \cos\alpha + 4d^2{\Phi_S}^2] . \quad (30)$$

The first term is only dependent on the exciting current I_{er} and corresponds to the cogging torque. The second term generates the motor torque to a decisive degree. The linear dependency of the rotor flooding $\Theta = I_{er}N$ and stator flux Φ_S can be seen here. The third term also generates a torque and is only dependent on the stator flux [8]. Operation of the motor with only the third term corresponds to operation of a reluctance motor.

An increase in the outer load torque leads to an increase in the load angle α and thus to a change in motor torque M_M to be delivered (Figure 22). The maximum motor torque to be delivered is referred to as the breakdown torque M_K at the position α_K. If α_K is exceeded, the machine slips out.

Alternatively, the torque can also be calculated approximately via the Lorentz force F_L [9]. This is the force which acts on moving charges in a magnetic field and is defined as

$$\vec{F}_L = I\left(\vec{l} \times \vec{B}\right) \quad (31)$$

Here, I is the current in the conductor, l the length of the conductor, and B the flux density of the effective magnetic field. Vectorial notation and the cross product mean that conductor and magnetic field must be perpendicular to each other in order to fashion a force.

By introducing an electric loading A which fashions the currents in the individual conductors of a machine at a fictitious line current density along the periphery, the Lorentz force

$$\vec{F}_L = A\left(\vec{S} \times \vec{B}\right) \tag{32}$$

can be described with the aid of the air-gap area S between the rotor and the stator. For the effective value of the electric loading the following applies

$$A = \frac{\hat{A}}{\sqrt{2}} \tag{33}$$

with sinusoidal flux distribution. The peak value of the electric loading is expressed with the number of stator phases m, the number of turns per phase N, and the pole pitch τ_p at the middle air-gap radius r_δ as

$$\hat{A} = \frac{m\sqrt{2}NI}{p\tau_p} = \frac{m\sqrt{2}NI}{\pi r_\delta}. \tag{34}$$

With the rotor length l in the axial direction the air-gap area S is obtained

$$S = 2\pi r_\delta l. \tag{35}$$

The mean value of a pole pitch is used for the flux density B. If the flux-density distribution along the periphery has a sinusoidal characteristics with the amplitude $\hat{B}$, then its mean value is given as

$$\begin{aligned} B &= \frac{1}{\frac{\pi}{p}} \int_0^{\frac{\pi}{p}} \hat{B}\sin(p\alpha)\,d\alpha \\ &= \frac{2}{\pi}\hat{B} \\ &\approx 0.64\hat{B}\,. \end{aligned} \tag{36}$$

With the relationship $M = r_\delta F_L$ the torque

$$M_M = 2mNIBlr_\delta \tag{37}$$

follows for a stationary operating state.

Suppression of harmonics

A sinusoidal flux distribution along the periphery of a stator is technically not feasible, therefore harmonics occur during operation in addition to the principal wave of the magnetic field. These lead to unwanted fluctuations of torque, vibrations, and additional losses. In order to specifically suppress the influence of certain magnetic-field harmonics, the rotor or stator slots are often skewed in electric machines. The skewing is intended to ensure that the same face of north and south poles of the harmonic to be neutralized are enclosed by the turns of a coil. The magnetic fluxes of these poles thus counterbalance each other and no longer impact negatively on the machine's operating characteristics.

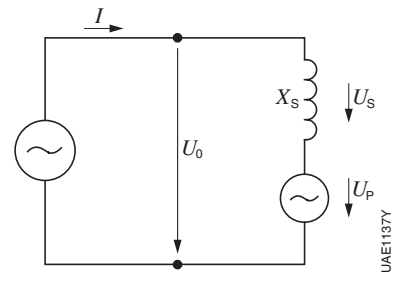
Another way of suppressing harmonics comes from the spatial distribution and chording of the coils. The effect of spatial distribution and chording is reflected in the winding factor. This is a measure of the coupling of windings with the respective magnetic flux-density waves.

Operating characteristics

A drawing of the synchronous machine can be made in a single-phase equivalent electric circuit diagram in that the voltage induced by the rotor in the stator (pole-wheel voltage U_P) is assumed as the voltage source and remaining reactances (inductive resistances) are summarized

Figure 23: Single-phase equivalent circuit diagram of the synchronous machine

U_P Pole-wheel voltage,
U_S Reactance voltage,
U_0 Terminal voltage,
I Current,
X_S Synchronous reactance.



to form the synchronous reactance X_S (Figure 23). The voltage above the synchronous reactance is designated U_S and the terminal voltage U_0. The direction of the current is specified in accordance with the consumer unit counting arrow system. Whereas current flows into the consumer units during motor operation, current flows out of the generator during generator operation. Setting up the mesh equation yields the current I

$$I = \frac{U_0 - U_P}{X_S}. \tag{38}$$

The amount of the pole-wheel voltage is influenced by the exciting current. The relationships are to be derived below. The following applies:

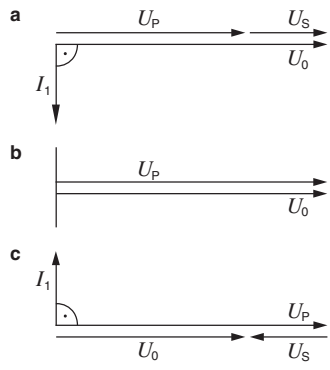
$$U_P = -\frac{d\Phi_R}{dt}. \tag{39}$$

With the cosinoidal flux Φ_R and the relationship

$$\Phi_R = BA_S \tag{40}$$

including its temporal derivation, this leads to

Figure 24: Operating states of the synchronous machine at idle
a) Underexcitation (inductive),
b) $I = 0$ (resistive),
c) Overexcitation (capacitive).
U_P Pole-wheel voltage,
U_S Reactance voltage,
U_0 Terminal voltage, I_1 Stator current.



$$\begin{aligned} u_P &= \Phi_R \omega_{\Phi S} \sin(\omega_{\Phi S} t) \\ &= B_R A_S \omega_{\Phi S} \sin(\omega_{\Phi S} t) \\ &= \mu H_R A_S \omega_{\Phi S} \sin(\omega_{\Phi S} t). \end{aligned} \tag{41}$$

The magnetic field intensity created in the rotor is described by Ampère's law. The pole-wheel voltage

$$\begin{aligned} u_P &= \mu \frac{\Theta_R}{2\delta} A_S \omega_{\Phi S} \sin(\omega_{\Phi S} t) \\ &= I_{er} \frac{\mu N}{2\delta} A_S \omega_{\Phi S} \sin(\omega_{\Phi S} t) \\ &= \hat{u}_P \sin(\omega_{\Phi S} t) \end{aligned} \tag{42}$$

is then linear-dependent on the exciting current I_{er}. The temporally changeable pole-wheel voltage is converted into its effective value using

$$U_P = \frac{\hat{u}_P}{\sqrt{2}}. \tag{43}$$

On the basis of the mesh equation (38), three operating states of the synchronous machine can be derived depending on the pole-wheel voltage (Figure 24):
Case 1: $U_P < U_0$, underexcitation, inductive,
Case 2: $U_P = U_0$, resistive,
Case 3: $U_P > U_0$, overexcitation, capacitive.

The first case occurs as long as U_P is $< U_0$. If $I_{er} = 0$, only the self-induction voltage is effective as induced voltage. If current is applied to the rotor, the mutual induction caused by the rotor also takes effect. The first case is referred to as underexcitation. The current lags behind the voltage by 90° ($\varphi(I,U) < 0$). The synchronous machine shows inductive characteristics.

Another increase in the exciting current leads to $U_P = U_0$. This produces the second operating case (Figure 24b). The current I_1 becomes zero if no more voltage is supplied via the synchronous reactance.

Another increase in the rotor current leads with $U_P > U_0$ to the third operating case (overexcitation).

All three cases apply to motor and generator operation. For the single-phase equivalent circuit diagram, the pointers are applied to the voltages and currents. Furthermore, the load angle β that sets

in between the voltages U_0 and U_S is defined. For motor operation, the load angle β is < 0 (Figure 25a). The voltage triangle is closed by the voltage U_S.

The synchronous reactance means that the current I_1 leading by 90° in relation to the voltage U_S flows. This is broken down into its components: the active current I_W and reactive current I_B (Figure 25a).

If the pole-wheel voltage is reduced so that the pointer of the reactance voltage is vertically on the pointer of the terminal voltage U_0, the motor only consumes active current (Figure 25b).

A further reduction in the pole-wheel voltage leads to underexcitation. The current I_1 lags behind the voltage U_S by 90°, which is the equivalent of inductive characteristics of the motor (Figure 25c). If torque is applied to the motor, it switches to generator operation. Generator operation features the positive load angle β (Figure 26). The sign of the current becomes negative. Current flows out of the machine. In the case of overexcitation, the machine behaves like a capacitor. It delivers reactive power (Figure 26a).

If the pole-wheel voltage is reduced so that the pointer of the reactance voltage U_S is vertically on the pointer of the terminal voltage, the generator only delivers active current (Figure 26b).

A further reduction in the pole-wheel voltage leads to a case of underexcitation. The machine behaves inductively. It consumes reactive power (Figure 26c).

Figure 25: Operating characteristics of the synchronous machine in motor operation
a) Overexcitation, b) Motor operation with consumption of active current, c) Underexcitation.
U_0 Terminal voltage, U_S Reactance voltage, U_P Pole-wheel voltage,
I_1 Current, I_W Active current, I_B Reactive current, β Load angle.

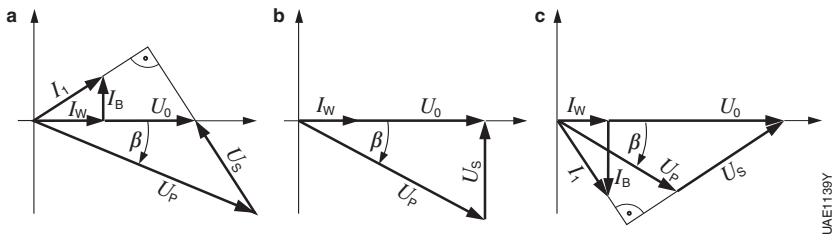
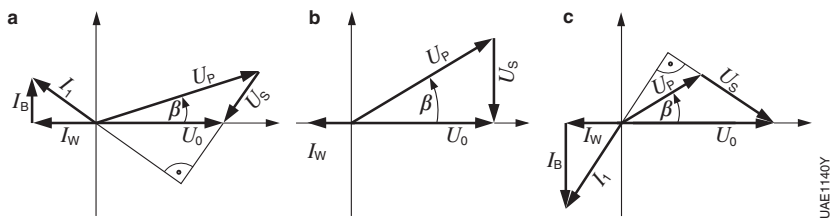


Figure 26: Operating characteristics of the synchronous machine in generator operation
a) Overexcitation (capacitive), b) Operation with active current delivery, c) Underexcitation (inductive).
U_0 Terminal voltage, U_S Reactance voltage, U_P Pole-wheel voltage,
I_1 Current, I_W Active current, I_B Reactive current, β Load angle.



Starting synchronous motors
Synchronous motors then only generate a permanently positive torque when the rotor and rotating-field speeds match. If this is not the case, the rotor's magnet poles are alternately attracted and repelled. The rotor starts to oscillate and also heats up strongly since the current consumption is not limited by the induced pole-wheel voltage. In the worst case the insulation of the windings is destroyed by the thermal load. To stop this from happening, the machine rotor must first be accelerated to synchronous speed. Three procedures have proven successful in the past for this purpose.

Procedure 1
The rotor is accelerated with the aid of a DC motor to synchronous speed before being put into operation.

Procedure 2
In salient-pole rotors a cage winding, also called a damper winding, can be integrated in the pole shoe. A structure of this type is pictured in Figure 27. This winding delivers in asynchronous mode the necessary torque to accelerate to synchronous speed. The windings are additionally short-circuited and deliver a further small torque contribution during running-up. If the machine is already at rotational speed, no further currents are induced on account of the absence of relative speed between rotor and stator speeds.

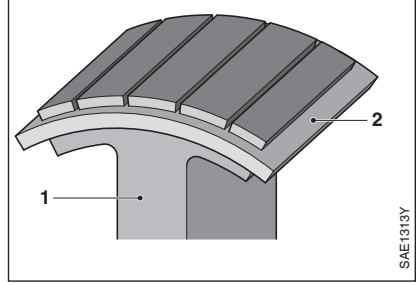
Procedure 3
The rotating-field speed is adapted with frequency converters to the rotor speed and slowly increased up to synchronous speed.

Switched reluctance motor

Properties of the switched reluctance motor
The "switched reluctance motor" (SRM) belongs to the class of synchronous motors. Its mode of operation is based on the effect of force on a boundary layer (reluctance force). Its design is comparable with many parallel-switched magnetic circuits which are continuously cyclically energized. Cyclic energization (switching with the necessary ECU) of stator-magnet teeth makes possible a rotary motion whereby the rotor teeth are brought into line with the stator teeth. Since rare-earth materials do not have to be used, areas of application open up in which motors with permanent-magnet material cannot be used. Areas of application at high temperatures are mentioned here by way of example.

The switched reluctance motor is suitable for variable-speed drive tasks and for positioning tasks. Thanks to its design, it can generate a holding torque over an extended period of time. It also generates high torque fluctuations.

Figure 27: Salient pole with integrated cage winding
1 Salient pole,
2 Cage winding.



Design of the switched reluctance motor

Figure 28 shows a cross-section of a four-phase ($m = 4$) switched reluctance motor. In the snapshot of Figure 28 the stator windings are energized (marked in dark gray in the figure). The current drives a magnetic flux through the four-pole arrangement of teeth A1, A2, A3, and A4. Since the rotor and stator teeth do not line up, a reluctance force is generated at the rotor-tooth surfaces whose tangential component has a torque-producing effect and can thus turn the rotor clockwise. The torque becomes zero as soon as the rotor and stator teeth line up. It is necessary to switch to further windings at the latest at the onset of this condition (see also [7], [10], [13]).

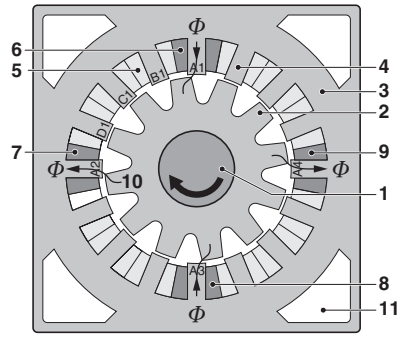
Calculating the number of stator and rotor teeth

A magnetic circuit provided with a phase winding contains a ferromagnetic flux concentrator (yoke and armature) and always an air gap for energy conversion. Two series-connected air gaps are required in the same magnetic circuit to facilitate a rotation of the armature. The magnetic flux penetrates the air gap once from the rotor to the stator and back from the stator to the rotor to complete the circuit. This is combined in a pole pair of the reluctance motor $Z_p = 1$. This also means that in this example $2Z_p$ is equal to the number of pole pairs per phase. The number of stator teeth N_S will be inferred by means of consideration whereby two rotor/stator teeth pairs in line (rotor tooth facing stator tooth) per phase form a magnetic circuit with the lowest reluctance (greatest inductance). Multiplying the number of poles per phase by the number of phases m produces the number of stator teeth $N_S = 2mZ_p$. To obtain a relative movement of the rotor in relation to the stator, it is necessary to achieve asymmetry of the numbers of teeth of stator and rotor. A practicable formulation for calculating the number of rotor teeth N_R is

$$\begin{aligned} N_R &= N_S \pm \Delta N \\ &= m \cdot 2\,Z_p \pm 1 \cdot 2Z_p \\ &= 2\,Z_p(m \pm 1)\,. \end{aligned} \tag{44}$$

The number of rotor teeth differs here precisely by an in-line tooth pair (pole pair) of a single phase (difference ΔN). Because of the magnetic flux between rotor and stator teeth, $N_R = 2Z_p(m+1)$ is preferred for an external rotor and $N_R = 2Z_p(m-1)$ is preferred for an internal rotor. For example, in Figure 28 $N_S = 2Z_p m = 2 \cdot 2 \cdot 4 = 16$ stator teeth and $N_R = 2Z_p(m-1) = 2 \cdot 2 \cdot (4-1) = 12$ rotor teeth can be seen.

Figure 28: Design of a switched reluctance motor as internal rotor
1 Rotor shaft, 2 Rotor tooth,
3 Stator, 4 Stator teeth,
5 Non-energized stator winding,
6, 7, 8, 9 Energized stator windings,
10 Flux Φ,
11 Cooling channel.



Calculating the step angles

Calculation of the step angle applies to the controlled stepping motor. The difference between stator angle α_S and rotor angle α_R where

$$\alpha_S = \frac{2\pi}{N_S} = \frac{2\pi}{m \cdot 2Z_p} \text{ and}$$

$$\alpha_R = \frac{2\pi}{N_R} = \frac{2\pi}{2Z_p(m \pm 1)} \tag{45}$$

yields the step angle θ where

$$\theta = \alpha_S - \alpha_R$$

$$= \frac{2\pi}{m \cdot 2Z_p} - \frac{2\pi}{2Z_p(m \pm 1)}$$

$$= \frac{\pm\pi}{Z_p m(m \pm 1)}$$

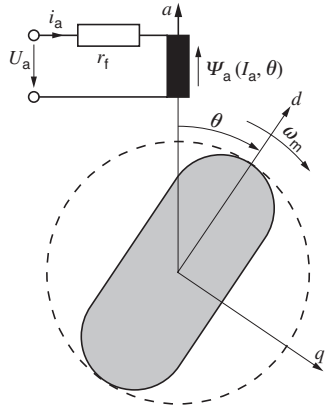
$$\theta = \frac{-\pi}{Z_p m(m-1)} = \frac{-\pi}{2 \cdot 4(4-1)} = -\frac{\pi}{24}. \quad (46)$$

Changing the step angle θ after the time t produces the angular velocity ω_m, from which the rotational frequency f_r is calculated. For example, for the stepping motor from Figure 28 a step angle $\theta = -\pi/24$ (see equation (46)) with a rotational frequency f_r is obtained:

$$f_r = \frac{1}{2\pi} \frac{d\theta}{dt}. \quad (47)$$

The rotor runs opposite to the phase switching direction.

Figure 29: Schematic representation of the switched reluctance motor
d Direct axis, q Quadrature axis,
a Strand (phase),
ω_m Rotor angular velocity,
θ Step angle,
Ψ_a Flux linkage of stator winding,
U_a Phase voltage, I_a Phase current,
r_f DC resistance of excitation winding.



Single-phase equivalent circuit diagram of the switched reluctance motor

The equivalent circuit diagram of the switched reluctance motor reduced to a phase a is depicted in Figure 29. The symbols and designations have been taken from VDI/VDE 3680 [10].

Operating characteristics of the switched reluctance motor

Voltage relationship

The operating characteristics are presented using the example of the voltage-differential and torque equation. The voltage-differential equation of the nonlinear magnetic circuit

$$u_a(t) = u_R(t) + u_L(t)$$

$$= i_a(t) r_f + \frac{d\Psi_a(I_a, \theta)}{dt} \quad (48)$$

is set up with the aid of Figure 29. By forming the total differential of $\Psi_a(I_a, \theta)$ and inserting in the voltage-differential equation it follows that

$$u_a(t) = i_a(t) r_f + \frac{\partial \Psi_a}{\partial I} \frac{di_a}{dt} + \frac{\partial \Psi_a}{\partial \theta} \frac{d\theta}{dt}. \quad (49)$$

To emphasize the angle dependence of the inductance, there follows with $\Psi_a = L_r(\theta) I_a$ the first-order non-homogeneous voltage-differential equation

$$u_a(t) = L_r(\theta) \frac{di_a}{dt} + \left(r_f + \frac{\partial L_r(\theta)}{\partial \theta} \frac{d\theta}{dt} \right) i_a. \quad (50)$$

It is clear that the induced voltage exhibits a proportionality to the rotor angular velocity. As the angular velocity increases the induced voltage rises, which has a current-lowering and thus torque-reducing effect, as is shown in the following.

Torque relationship

With the aid of the angle-dependent co-energy W_{mag}^{Co} of the linear (Figure 30a) and nonlinear magnetic circuits (Figure 30b) the mechanical energy W_{mech} follows where

$$\Delta W_{mag}^{Co} = W_{mag}^{Co}(\theta_1) - W_{mag}^{Co}(\theta_2)$$

$$= W_{mech}. \quad (51)$$

The magnetic co-energy W_{mag}^{Co} corresponds to the area (energy) below the respective characteristic.

The co-energy together with the magnetic energy always produce a rectangle or a square in the diagram. With the restriction to the linear magnetic circuit according to Figure 30a it follows that

$$\Delta W_{mag}^{Co} = \frac{1}{2} I_a^2 \left(L(\theta_1) - L(\theta_1) \right)$$

$$= W_{mech} .$$

What is clear is the quadratic current function, which decisively influences the area W_{mech}. The larger this area, the greater the rated torque m_A,

$$m_A = \frac{\partial W_{mech}}{\partial \theta},$$

which is calculated with the angle-dependent co-energy. In the angle position θ_1 the rotor and stator teeth face each other in line. In this angle position the minimum reluctance and consequently the maximum interlinked flux Ψ_{max} are obtained. The torque equals zero in this position. A torque is obtained in the angle position θ_2. The stator tooth faces the rotor grove. The reluctance value assumes a maximum, the flux Ψ a minimum.

The symbols and abbreviations used have been taken from VDI/VDE 3680 [10].

Electronic motors

In the case of electronically commutating motors (electronic motor, EC motor), the rotor excitation winding, including the electrical contacting with collector rings, is not required. Electronic motors are brushless synchronous motors where the rotors are fitted with permanent magnets. The permanent magnets can be arranged, for example, on the rotor surface or in the rotor (Figure 31). The commutation of the current takes place generally in the fixed stator winding by means of an electronic assembly (Figure 33).

Figure 31: Variations of rotors for the electronic motor
a) Rotor with surface magnets,
b) Rotor with embedded magnets (buried magnets).
Φ Rotor flux.

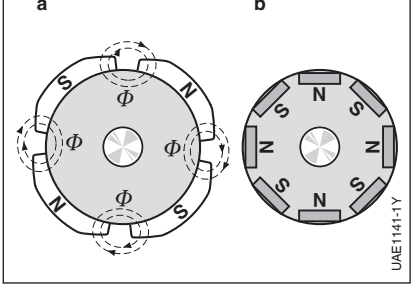
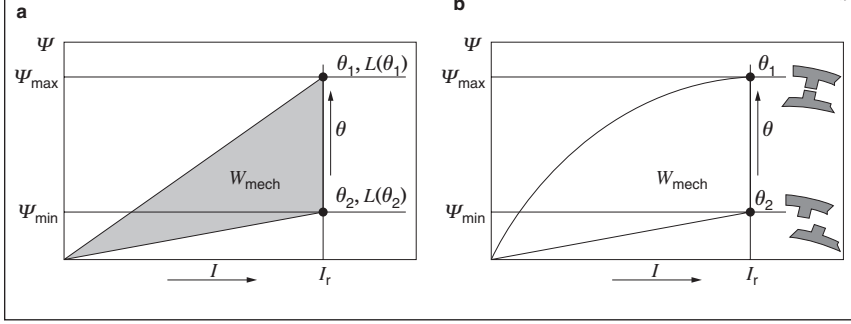
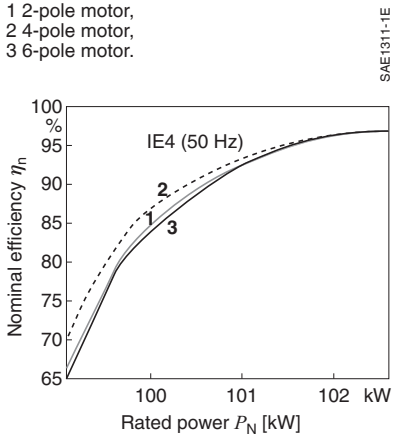


Figure 30: Torque-producing mechanical energy
a) Linear magnetic circuit,
b) Nonlinear magnetic circuit.



The rotational speed of the electronic motor is set by the frequency of the surrounding stator field. Sensors are required to pick up the rotor position. Widespread use is made of Hall sensors fitted in the working air gap to enable cyclic relaying between the winding phases with the help of activation electronics.

Figure 32: Motor efficiencies
1 2-pole motor,
2 4-pole motor,
3 6-pole motor.



Energy-efficiency classes

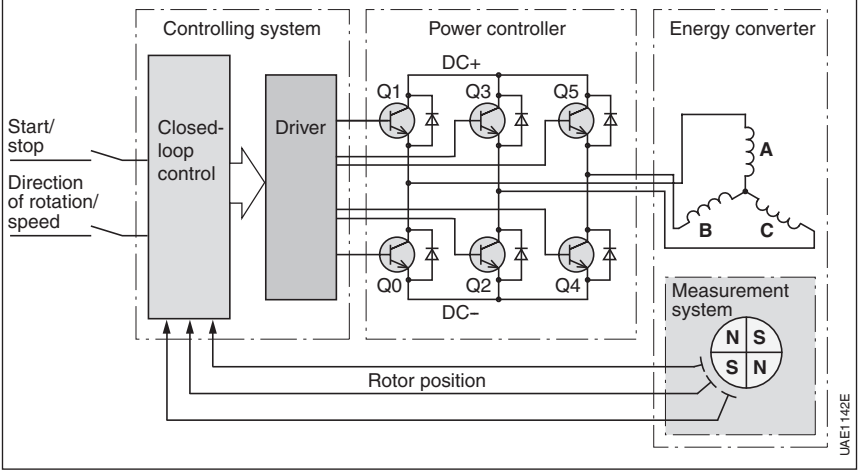
The International Electrotechnical Commission (IEC) intends with the standard IEC 60034 Part 30-1 [8] to harmonize internationally efficiency classes for electric motors. The standard applies to all types of electric motors which are rated for direct start-up at the power system. The efficiency classes are named in Table 3. In Figure 32 the motor efficiencies are plotted against the rated power for 2-, 4- and 6-pole motors at 50 Hz.

The area of application of the IE4 standard is presented in a shortened version:
– Rated power P_N:
  0.12 kW through 1000 kW;
– Rated voltage U_N:
  50 V through 1 kV;
– Number of poles from 2, 4, 6 or 8.

Table 3: Efficiency classes

Abbreviation	Name
IE1	Standard-Efficiency
IE2	High-Efficiency
IE3	Premium-Efficiency
IE4	Super-Premium-Efficiency

Figure 33: Activation electronics of the electronic motor



Three-phase current system

Technically relevant is the application of the three-phase alternating current system as a three-phase current system, the major feature of which is that the total of all voltages and currents is zero at all times.

Definitions
The electric circuits are referred to as phases m. The total of the electric circuits in which voltages of the same frequency have an effect and are phase-shifted are referred to as multiphase systems. A multiphase system consists of winding phases. On a three-phase system, $n = 3$ symmetrical systems are possible (Figure 34). On all symmetrical systems – with the exception of the zero system – the total of all pointers is zero. m phases result in n different symmetrical systems depending on the phase difference α:

$$\alpha = \frac{2\pi\, n}{m} . \qquad (52)$$

The task of windings is to generate a rotating field. Asynchronous and synchronous machines have the same stator structure. In the air gap, a magnetic field with constant amplitude is to be created, rotating at constant angular velocity. For this rotating field to be generated, the temporal phase positions of the currents must match the spatial position of the corresponding strands. For a simple symmetrical system ($n = 1$) with $m = 3$, the three strands (designated with U, V and W) and therefore the windings must be evenly distributed over the circumference. Figure 35 shows the arrangement of a three-strand winding with one coil per pair of poles and strand The terminal markings of the phases comply with DIN EN 60034, Part 8 [8].

Figure 34: Symmetrical systems
a) Co-system, $n = 1$, $\alpha = 2\pi/3$ (120°),
b) Counter system, $n = 2$, $\alpha = 4\pi/3$ (240°),
c) Zero system, $n = 1$, $\alpha = 0$.
n Symmetrical systems,
α Phase difference.

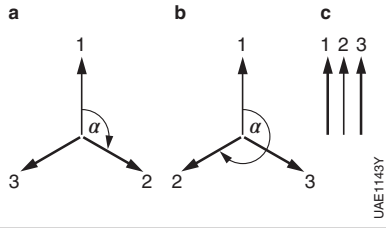


Figure 35: Winding of a two-pole motor with one pair of poles per strand
a) Pole arrangement,
b) Internal circuitry.
1 Stator, 2 Rotor.
U, V, W Strands.

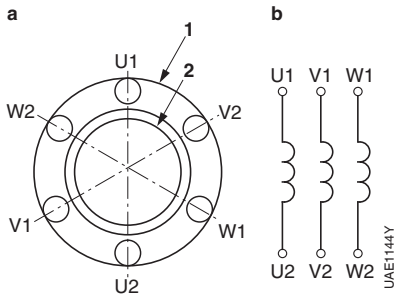
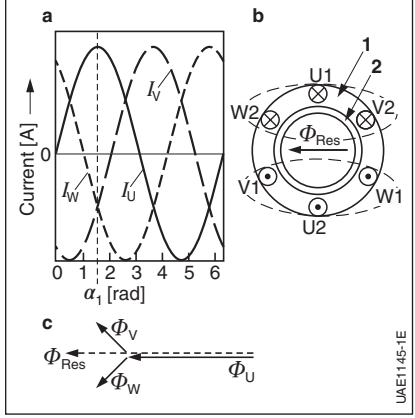


Figure 36: Generating a rotating field with one coil per strand
a) Strand currents,
b) Strand currents with angle $\alpha = \alpha_1$,
c) Direction of the rotating field (spatial directions).
1 Stator, 2 Rotor.
U, V, W Strands.



Rotating-field generation

In order to generate a rotating field in the case of a simple symmetrical system ($n = 1$) with the number of strands $m = 3$, the strands must be geometrically offset by the electrically effective angle

$$\alpha_{el} = 360° \cdot \frac{1}{3} = 120°.$$

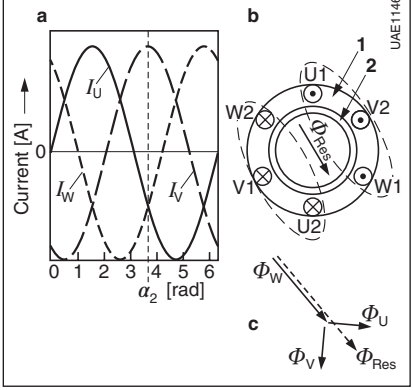
With one coil per pair of poles and strand, the resulting magnetic field rotates counterclockwise in that the "indicator bar" shifting in Figure 36a to the right (at $\alpha = 90°$) indicates the phase current in each of the strands in Figure 36b in flux direction. The arrangement forms a pair of poles. The associated magnetic fluxes emerge vertically in relation to the winding plane of the strands (Figure 36b).

The flux Φ_{Res} (Figure 36c) resulting from the three strands as well as its direction are achieved by geometrically adding up the three individual fluxes Φ_U, Φ_V and Φ_W.

A progression of the indicator bar up to $\alpha = 180°$ leads to a current-direction reversal in strand W and therefore to a further turn to the right of the resulting field Φ_{Res} (Figure 37).

Figure 37: Generating a rotating field with one coil per strand

a) Strand currents,
b) Strand currents with $\alpha = \alpha_2$,
c) Direction of the rotating field (spatial directions).
1 Stator, 2 Rotor.
U, V, W Strands.



When two coils per strand are used, the conductor arrangement is "doubled". If the winding is to form two pairs of poles ($p = 2$), the windings have to be divided into groups (Figure 38). With this, the mechanically effective angle

$$\alpha_m = 360° \cdot \frac{1}{m\,p} = 60° \text{ sets in.}$$

Figure 38: Winding with two pairs of poles per strand

a) Pole arrangements,
b) Example of the internal wiring of the four-pole motor with two windings per pole pair and strand.
The interconnections in parentheses are internal and thus inaccessible.
1 Stator, 2 Rotor.
U, V, W Strands.

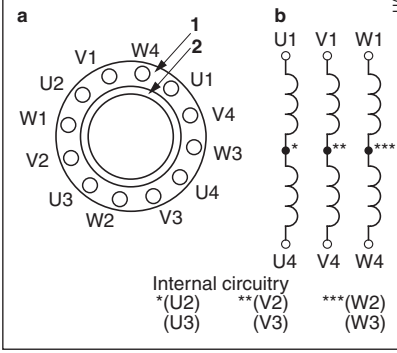
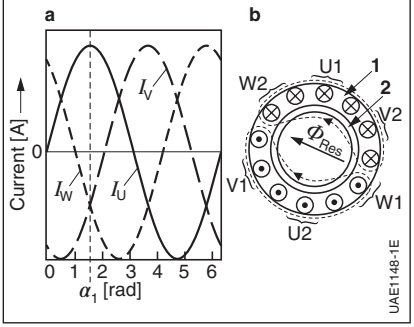


Figure 39: Rotating-field generation with two coils per strand

a) Strand currents with angle $\alpha = \alpha_1$,
b) Resulting magnetic field with $\alpha = \alpha_1$.
1 Stator, 2 Rotor.
U, V, W Strands.



The electrically effective angle remains unchanged. In the case of both the two-pole and four-pole arrangement, the field rotates counterclockwise (Figure 39). The rotating-field speed

$$n_d = \frac{f_n}{p}$$

can be calculated with the mains frequency f_n and the number of pole pairs p. For p = 1, the rotating-field speed is equal to the mains frequency (Table 4).

Together with the number of pole pairs, the pole pitch

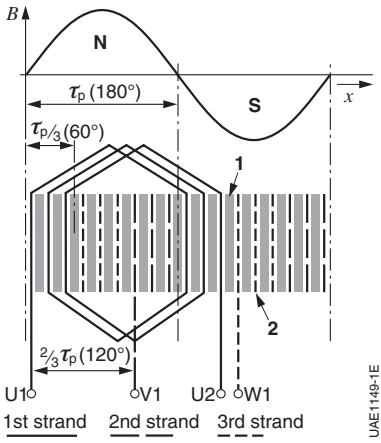
$$\tau_p = \pi d_{si}/2p$$

can be calculated as a stator circumference proportion, where d_{si} is the inner diameter of the stator. It corresponds to the length of a sinusoidal half wave that corresponds to the induction distribution of the rotor field. In the case of a two-pole machine (p=1), the pole pitch is α_{el}=180° (electrical angle) at all times and matches the mechanical angle α_m. The relationship of the two angles is indicated by $\alpha_{el} = p\,\alpha_m$. So that voltages of equal size are induced in the windings, the winding phases have to be offset in relation to one another by α_{el} = 120° or $2\tau_p/3$ and the coil structure and number of turns must be the same. One third of the pole pitch is apportioned to each strand.

Table 4: Rotating-field speeds

Pair of poles p	n_0 [rpm] at f=50 Hz
1	3000
2	1500
3	1000

Figure 40: Structure of a stator three-phase current winding
1 Stator tooth,
2 Stator slot with winding phase.



References
[1] R. Fischer: Elektrische Maschinen, 13th Edition, Carl Hanser Verlag, 2006.
[2] K. Fuest, P. Döring: Elektrische Maschinen und Antriebe, 6th Edition, Vieweg-Verlag, 2008.
[3] DIN 42027: Servo motors; classification, survey.
[4] DIN EN 60027: Letter symbols to be used in electric technology – Part 4: Rotating electric machines.
[5] DIN EN 60034, Part 8: Rotating electrical machines; terminal markings and direction of rotation.
[6] DIN EN 60617: Graphical symbols for diagrams – Part 6: Production and conversion of electrical energy.
[7] DIN 1304, Part 7: Letter symbols for electrical machines.
[8] DIN EN 60034, Part 30-1: Rotating electrical machines; efficiency classes of line operated AC motors.
[9] Gieras, J. F.; Wang, R. J.; Kamper, M. J.: Axial Flux Permanent Magnet Brushless Machines; Springer Verlag.
[10] VDI/VDE 3680: Control of synchronous machines.
[11] I. Wolff: Maxwellsche Theorie – Grundlagen und Anwendungen; Volume 1, Elektrostatik, 5th Edition, Verlagsbuchhandlung Dr. Wolff, 2005.
[12] I. Wolff: Maxwellsche Theorie – Grundlagen und Anwendungen; Volume 2, Strömungsfelder, Magnetfelder und Wellenfelder, 5th Edition, Verlagsbuchhandlung Dr. Wolff, 2007.
[14] Binder, A.: Elektrische Maschinen und Antriebe – Grundlagen, Betriebsverhalten; Springer Verlag.
[16] McPherson, G. Laramore, R. D.: An Introduction to Electrical Machines and Transformers; Wiley Verlag.

Chemistry

Elements

Periodic table

Structure of the periodic table

The atoms of the chemical elements are made up of positively charged protons, uncharged neutrons, and negatively charged electrons [1]. In the periodic table (Tables 1 and 2) the elements are laid out in order of increasing number of protons – i.e. increasing nuclear charge number – and in order of their atomic mass. The atomic mass is essentially determined by the total number of nuclear constituents, i.e. the sum total of protons and neutrons. The total number of protons and neutrons is also called the mass number. The number of protons corresponds to the atomic number. The neutral element atoms always have as many protons and electrons.

Groups and periods

The elements are divided in the periodic table into different groups (vertical columns) and periods (horizontal rows). The nested structure of the groups is due to the fact that the electrons always occupy the lowest energy levels. The position of these energy levels, which are also called electron orbitals and indicate the probability of electrons being present around the atomic nucleus, can be derived using quantum mechanics.

Quantum numbers

The structure of the periodic table is based on four quantum numbers (principal, secondary, magnetic, and spin quantum numbers), with which the four electron orbitals with the designations s, p, d and f can be calculated. When the electrons are arranged into these orbitals in order of increasing energy, this creates groups of elements which demonstrate a similar manner of reaction across all the periods. This manner of reaction is only minimally influenced by the electrons on the inner orbitals. The crucial factors are the energy and number of outer electrons. The outer electrons are frequently referred to as “valence electrons”.

Main groups

The elements in groups Ia, IIa and IIIa…VIIIa are called main-group elements. The main-group elements include hydrogen and the alkali metals (Ia), the alkaline-earth metals (IIa), the boron (IIIa), carbon (IVa) and nitrogen groups (Va), the chalcogens (VIa), the halogens (VIIa), and the noble gases (VIIIa). The electrons of the main-group elements of the 1st period, hydrogen and helium, are found exclusively in s-orbitals. In the case of the other main-group elements in the 2nd through 7th periods, the electrons also occupy p-orbitals starting from the main group IIIa.

Secondary groups

The elements of the secondary groups Ib, IIb and IIIb…VIIIb with electrons in d-orbitals all have a metallic character. The copper group is assigned to the secondary group Ib because its electron configuration shows similarities to the main group Ia. The elements in both groups tend to form salts from monovalent ions. The same applies to the groups IIa and IIb. Divalent metal compounds are formed in the group of alkaline-earth metals and the zinc group IIb. The designations for the secondary groups IIIb…VIIb likewise have their roots in the electron configuration and provide an indication of the maximum valence of these metal ions. The elements iron, cobalt and nickel as well as the elements underneath called higher homologs are brought together as secondary group VIIIb due to their marked chemical similarity.

Table 1: Periodic table of elements

Ia												IIIa	IVa	Va	VIa	VIIa	VIIIa
1 **H** 1.008	IIa							VIIIb			Ib	IIb						2 **He** 4.003
3 **Li** 6.941	4 **Be** 9.012												5 **B** 10.811	6 **C** 12.011	7 **N** 14.007	8 **O** 15.999	9 **F** 18.998	10 **Ne** 20.180
11 **Na** 22.990	12 **Mg** 24.305	IIIb	IVb	Vb	VIb	VIIb							13 **Al** 26.982	14 **Si** 28.086	15 **P** 30.974	16 **S** 32.066	17 **Cl** 35.453	18 **Ar** 39.948
19 **K** 39.098	20 **Ca** 40.078	21 **Sc** 44.956	22 **Ti** 47.87	23 **V** 50.942	24 **Cr** 51.996	25 **Mn** 54.938	26 **Fe** 55.845	27 **Co** 58.933	28 **Ni** 58.693	29 **Cu** 63.546	30 **Zn** 65.39	31 **Ga** 69.723	32 **Ge** 72.61	33 **As** 74.922	34 **Se** 78.96	35 **Br** 79.904	36 **Kr** 83.80
37 **Rb** 85.468	38 **Sr** 87.62	39 **Y** 88.906	40 **Zr** 91.224	41 **Nb** 92.906	42 **Mo** 95.94	43 **Tc** (98)	44 **Ru** 101.07	45 **Rh** 102.906	46 **Pd** 106.42	47 **Ag** 107.868	48 **Cd** 112.411	49 **In** 114.818	50 **Sn** 118.710	51 **Sb** 121.760	52 **Te** 127.60	53 **I** 126.904	54 **Xe** 131.29
55 **Cs** 132.905	56 **Ba** 137.327	57 **La*** 138.906	72 **Hf** 178.49	73 **Ta** 180.948	74 **W** 183.84	75 **Re** 186.207	76 **Os** 190.23	77 **Ir** 192.217	78 **Pt** 195.078	79 **Au** 196.967	80 **Hg** 200.59	81 **Tl** 204.383	82 **Pb** 207.2	83 **Bi** 208.980	84 **Po** (209)	85 **At** (210)	86 **Rn** (222)
87 **Fr** (223)	88 **Ra** (226)	89 **Ac**** (227)	104 **Rf** (267)	105 **Db** (268)	106 **Sg** (271)	107 **Bh** (267)	108 **Hs** (277)	109 **Mt** (274)	110 **Ds** (282)	111 **Rg** (280)	112 **Cn** (285)	113 **Nh** (284)	114 **Fl** (289)	115 **Mc** (291)	116 **Lv** (293)	117 **Ts** (292)	118 **Og** (294)

*	58 **Ce** 140.116	59 **Pr** 140.908	60 **Nd** 144.24	61 **Pm** (145)	62 **Sm** 150.36	63 **Eu** 151.964	64 **Gd** 157.25	65 **Tb** 158.925	66 **Dy** 162.50	67 **Ho** 164.930	68 **Er** 167.26	69 **Tm** 168.934	70 **Yb** 173.04	71 **Lu** 174.967
**	90 **Th** 232.038	91 **Pa** 231.036	92 **U** 238.029	93 **Np** (237)	94 **Pu** (244)	95 **Am** (243)	96 **Cm** (247)	97 **Bk** (247)	98 **Cf** (252)	99 **Es** (252)	100 **Fm** (257)	101 **Md** (258)	102 **No** (259)	103 **Lr** (262)

All elements are arranged sequentially according to atomic number (proton number). The horizontal rows are called periods, the vertical columns are called groups. The relative atomic masses are indicated below the element symbols.
The values given in parentheses are the mass numbers (nucleon numbers) of the stablest isotopes of radioactive elements.

Table 2: Designations of chemical elements

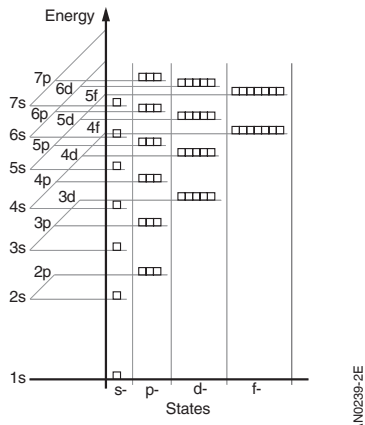
Element	Symbol	Atomic number
Actinium	Ac	89
Aluminum	Al	13
Americium [1]	Am	95
Antimony	Sb	51
Argon	Ar	18
Arsenic	As	33
Astatine	At	85
Barium	Ba	56
Berkelium [1]	Bk	97
Beryllium	Be	4
Bismuth	Bi	83
Bohrium [1]	Bh	107
Boron	B	5
Bromine	Br	35
Cadmium	Cd	48
Carbon	C	6
Cesium	Cs	55
Calcium	Ca	20
Californium [1]	Cf	98
Cerium	Ce	58
Chlorine	Cl	17
Chromium	Cr	24
Copernicium [1]	Cn	112
Copper	Cu	29
Cobalt	Co	27
Curium [1]	Cm	96
Darmstadtium [1]	Ds	110
Dubnium [1]	Db	105
Dysprosium	Dy	66
Einsteinium [1]	Es	99
Erbium	Er	68
Europium	Eu	63
Fermium [1]	Fm	100
Fluorine	F	9
Flerovium [1]	Fl	114
Francium	Fr	87
Gadolinium	Gd	64
Gallium	Ga	31
Germanium	Ge	32
Gold	Au	79
Hafnium	Hf	72
Hassium [1]	Hs	108
Helium	He	2
Holmium	Ho	67
Hydrogen	H	1
Indium	In	49
Iodine	I	53
Iridium	Ir	77
Iron	Fe	26
Krypton	Kr	36
Lanthanum	La	57
Lawrencium [1]	Lr	103
Lead	Pb	82
Lithium	Li	3
Livermorium [1]	Lv	116
Lutetium	Lu	71
Magnesium	Mg	12
Manganese	Mn	25
Meitnerium	Mt	109
Mendelevium [1]	Md	101
Mercury	Hg	80
Molybdenum	Mo	42
Moscovium [1]	Mc	115
Neodymium	Nd	60
Neon	Ne	10
Neptunium [1]	Np	93
Nickel	Ni	28
Nihonium [1]	Nh	113
Niobium	Nb	41
Nitrogen	N	7
Nobelium [1]	No	102
Oganesson [1]	Og	118
Osmium	Os	76
Oxygen	O	8
Palladium	Pd	46
Phosphorus	P	15
Platinum	Pt	78
Plutonium [1]	Pu	94
Polonium	Po	84
Potassium	K	19
Praseodymium	Pr	59
Promethium	Pm	61
Protactinium	Pa	91
Radium	Ra	88
Radon	Rn	86
Rhenium	Re	75
Rhodium	Rh	45
Roentgenium [1]	Rg	111
Rubidium	Rb	37
Ruthenium	Ru	44
Rutherfordium [1]	Rf	104
Samarium	Sm	62
Scandium	Sc	21
Seaborgium [1]	Sg	106
Selenium	Se	34
Silicon	Si	14
Silver	Ag	47
Sodium	Na	11
Strontium	Sr	38
Sulfur	S	16
Tantalum	Ta	73
Technetium	Tc	43
Tellurium	Te	52
Tennessine [1]	Ts	117
Terbium	Tb	65
Thallium	Tl	81
Thorium	Th	90
Thulium	Tm	69
Tin	Sn	50
Titanium	Ti	22
Tungsten	W	74
Uranium	U	92
Vanadium	V	23
Xenon	Xe	54
Ytterbium	Yb	70
Yttrium	Y	39
Zinc	Zn	30
Zirconium	Zr	40

[1] Artificially produced; does not occur naturally.

Periods with f-orbitals
In the 6th and 7th periods a further series of energy levels, the f-orbitals, are available in each case after the secondary group IIIb lanthanum and actinium. These are taken up by the electrons of the elements of the lanthanides (6th period) and actinides (7th period). The lanthanides are also known by the term “rare earths”. All actinides are radioactive.

The arrangement of the elements in the periodic table according to the energy position of their electron orbitals becomes clear in the energy-level diagram (Figure 1). It can be seen that for higher orbital energies after the occupation of the s-orbital the p-, d- and f-orbitals subordinate to this principal quantum number are no longer automatically filled with electrons. For example, it is now understandable why the secondary-group elements scandium through zinc with the orbital energies 3d after the element calcium are placed into the 4th period of the periodic table. For higher nuclear charge numbers, however, the differences between the orbital energies are so small that the energy sequence of the orbitals shown in the energy-level diagram no longer applies to each individual element in the periodic table. In these cases the exact energy position is influenced by the partial, semi- or full occupation of the orbitals by electrons. This can be seen in the case of the lanthanides with the 4f-orbitals, which are only placed after the element lanthanum into the 6th period and not – as would be expected according to the energy-level diagram – already after the element barium.

Figure 1: Energy-level diagram of the electron orbitals



Isotopes
Isotopes are atoms of the same element with the same number of protons but with a different mass number, i.e. with a different number of neutrons. To give isotopes a unique designation, the mass number is given at top left next to the atom symbol and the number of protons is given at bottom left so that the number of neutrons can be calculated straight away by subtraction. Most naturally occurring elements are isotopic mixtures. Carbon, for example, is made up 98.89 % of ${}^{12}_{6}C$ and 1.11 % of ${}^{13}_{6}C$. The proportion of ${}^{14}_{6}C$ is at 10^{-10} % very low. ${}^{14}_{6}C$ in comparison with ${}^{12}_{6}C$ and ${}^{13}_{6}C$ not a stable isotope. The ratio of ${}^{14}_{6}C$ to ${}^{12}_{6}C$ is used in the radiocarbon method to determine the age of organic matter.

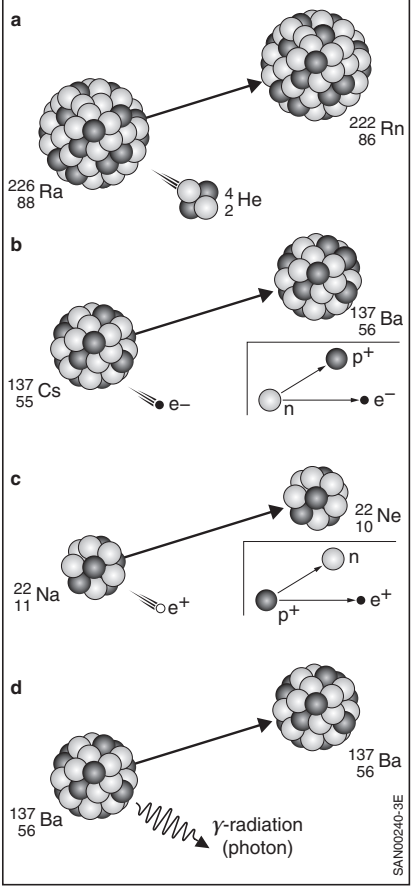
Nuclides
The term nuclide is broadly defined. All atoms that differ in the constitution of their nucleus are called nuclides. The number of nuclides therefore corresponds to the number atom types. The 8th Edition of the Karlsruhe Nuclide Chart in 2012 lists 3,847 experimentally verified nuclides and isomers. Isomers are nuclides which have the same mass number and thus the same number of protons and neutrons. Isomeric nuclides, however, show differences in the internal state of the atomic nuclei. Atomic nuclei can assume excited states as well as the normal state. Only around 270 nuclides are stable. The majority of nuclides are radioactive and are therefore called radionuclides.

Radioactive decay

α-decay

In the case of natural radioactivity [2] α-decay is observed when the atomic nucleus emits a double positively charged helium nucleus He^{2+} consisting of two protons and two neutrons. The mass number of the atomic nucleus is reduced by 4 and the nuclear charge number by 2. A different element is created (example, see Figure 2a).

Figure 2: Types of radiation for natural radioactive decay
a) α-radiation,
b) β^--radiation,
c) β^+-radiation,
d) γ-radiation,
n Neutron,
p Proton,
e^- Electron,
e^+ Positron.



β-decay

An element transmutation also occurs in the case of radioactive β-decay. β^- decay is observed when a neutron from the atomic nucleus emits an electron (example, see Figure 2b). This increases the number of protons by 1 and the element with the next higher atomic number is created. When instead of an electron a positron, i.e. a positively charged particle, is emitted, β^+-decay takes place. A proton is transformed into a neutron in the case of β^+-decay. Where the mass number is the same, the nuclear charge number is therefore decreased by 1 such that the preceding element in the periodic table is created (example, see Figure 2c). β-decay also witnesses the release of antineutrinos (in the case of β^--delay) or neutrinos (in the case of β^+-decay), which cannot be discussed further at this stage.

γ-decay

In the case of γ-decay the atomic nucleus emits internal energy from excited nucleus states which for the most part arise during α- or β-decay. The γ-radiation emitted during γ-decay does not change the mass number; the element remains intact (example, see Figure 2d).

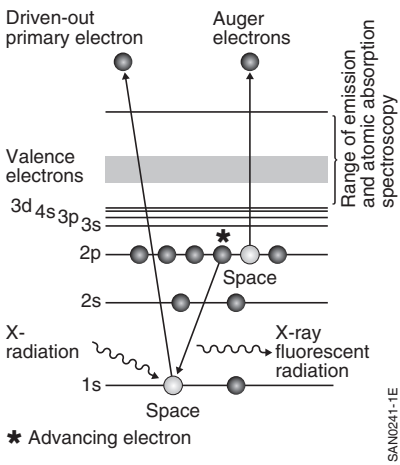
Artificial element transmutation

Elements can also be artificially transmuted into each other by bombardment with high-energy particles. Neutrons, singly or doubly positively charged hydrogen (H^+) and helium nuclei (He^{2+}), but also γ-radiation are suitable for this purpose. Depending on the type and acceleration of the particles, this results in the particle being absorbed in the atomic nucleus – sometimes also involving the emission of a proton or neutron – or in nuclear fission.

Half-life
Radioactive decay is a monomolecular reaction (first-order reaction) in which the decaying share per unit of time is directly proportional to the available quantity. Because the available quantity constantly diminishes due to the radioactive decay continuously taking place, the reaction rate decreases more and more. The half-life of the radionuclide is specified so that the rate of radioactive decay can be described independently of the quantity. The half-life is the period of time after which half the atomic nuclei have decayed. The shorter the half-life of the radioactive element, the higher its specific activity, i. e. the more nuclear decays occur per unit of mass.

Element spectroscopy
All chemical elements differ in their atomic structure and can therefore be identified by means of their electron spectra and quantified into mixtures [3]. To release electrons from their energy levels, it is necessary to supply energy to the order of their binding energy from an outside source.

Figure 3: Electron migrations in element spectroscopy



X-ray fluorescence analysis
The inner electrons more tightly bound by the attraction of the positive atomic nucleus can only be released with X-radiation or with electrons of sufficient energy. X-ray fluorescence analysis (XRF) is based on the fact that the spaces created on the more favorable inner energy levels are filled by higher-energy electrons. These advancing electrons release the excess energy in the form of fluorescent radiation. The energy of the fluorescent radiation corresponds exactly to the difference between the higher and the lower energy levels and is thus element-specific (Figure 3).

Auger electron spectroscopy
The energy released by advancing electrons can be output as fluorescent radiation, but can also be transmitted to other electrons. An electron can leave the atom with the transmitted energy. These electrons from the so-called Auger process likewise contain highly specific element information, since the following contributions determine the energy of the Auger electron: The orbital energy of the originally driven-out electron, the energy that is released by the advancing electron, and the energy of the electron emitted in the Auger process itself (Figure 3). The energy of the emitted electrons is analyzed in Auger electron spectroscopy.

Emission spectrometry
The less tightly bound outer valence electrons can already be thermally excited and thereby raised to higher, unoccupied energy levels. Because the energy from the flame of a gas torch is not sufficient to excite the valence electrons of all the elements, a plasma is often used for excitation. In the case of inductively coupled plasma optical emission spectrometry (ICP-OES), the light released during the fallback of the valence electrons that is in the visible and ultraviolet range is registered. As with the inner electrons, the energy difference between the outer occupied and unoccupied energy levels is also element-specific.

Atomic absorption spectroscopy
Instead of emitted radiation, light absorption can also be measured through the excitation of the valence electrons. To this end, in atomic absorption spectroscopy (AAS) the emitted radiation of the element to be analyzed is irradiated by a hollow-cathode lamp. When the element in the sample matches the element of the hollow cathode, the excitation of the electrons in the sample causes a weakening in the intensity of the excitation light. Because the verification of different elements accordingly requires an extensive range of materials for hollow-cathode lamps and can only be measured sequentially, atomic absorption spectroscopy has not established itself in multi-element analyses.

X-ray photoelectron spectroscopy
In X-ray photoelectron spectroscopy (XPS) electrons are driven out from all the energy levels in the atom by high-energy X-rays. The element-specific binding energy can be inferred from the kinetic energy of these electrons. Only the valence electrons are released by the ultraviolet light used instead of X-rays in ultraviolet photoelectron spectroscopy (UPS). The kinetic energy of valence electrons can be determined very precisely such that even conclusions can be made as to the type of chemical bonds by way of differences in the orbital energies.

X-ray diffraction
X-ray diffraction on solids is, as the name suggests, not a spectroscopic procedure. The X-radiation diffracted on the crystal lattice is registered with angular resolution, i.e. according to its diffraction angle [4]. The angle position, intensity and width of the diffraction peaks enable the lattice structure to be analyzed and thus crystals to be identified (crystal-structure analysis). Crystallite size, preferred orientations (textures) and lattice distortions can also be determined in mixtures. Internal mechanical stresses of components can likewise be determined.

Chemical bonds

The type of bond which tends to be adopted by the individual elements is determined by the number and arrangement of electrons around the atomic nucleus (electron configuration) [5]. The nature of these bonds can differ greatly. A distinction is made between ionic, covalent and metallic bonds. Single atoms without a chemical bond occur in nature in noble gases and in some elements in the vapor phase.

Additional, but far weaker interactions exist between molecules, i.e. particles which are composed of two or more atoms. The weak attractive forces, which have different physical causes, mean that not only ions but also molecules adopt a short-range order, which is only eliminated in the gaseous phase.

Ionic bonds
Ionic bonds tend to form in compounds between metallic and non-metallic elements. This results in the transfer of electrons. Metal atoms (electron donors) lose electrons and become positively charged cations. By gaining electrons, the non-metal atoms (electron acceptors) become negatively charged anions. The terms “cation” for a positively and “anion” for a negatively charged ion are so called because these ions move in an aqueous solution in the electric field to the oppositely charged electrodes, i.e. the cation to the negatively charged cathode and the anion to the positively charged anode.

The number of lost and gained electrons is determined by the electron configuration of the element. Ionic compounds with configurations in which the p-, d- and f-orbitals are empty, half-full and full are especially preferred.

The ionic compounds created are called salts. In a solid the cation and anion are arranged in an ionic lattice whose structure is determined by the ratio of ionic radii. When salts are melted or dissolved in water, the forces between the ions in the lattice must be overcome. Therefore heat of fusion (enthalpy of fusion) must be supplied. The energy required for dissolving (heat of solution) is determined by two opposing components. Firstly, energy to dissolve the crystal lattice must be applied (heat of dissociation); secondly, energy is released by coordination of the solvent molecules to the dissolved-out ions (heat of solvation). If the solvent is water, the heat is called heat of hydration. If the energy of dissociation is greater than the heat of solvation, the solution cools during the dissolving process. When anhydrous metal salts are dissolved in which initially water is also introduced into the ionic lattice, the heat released by hydration is often greater than the energy consumption to overcome the lattice forces, and the solution heats up.

Covalent bonds

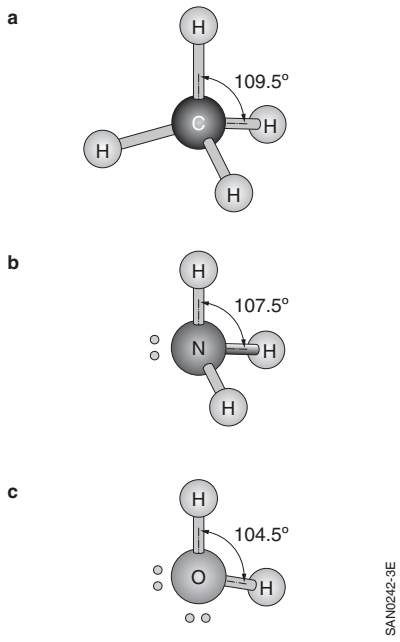
Bonds of electron pairs which are created between two neutral atoms from the unpaired electrons in the outer orbitals (valence electrons) are called atomic bonds. The term "covalent" illustrates that two identical or similar atoms can, by forming a common electron pair, achieve an energetically more favorable electron configuration without involving an electron transfer and with it a change in the valency by the formation of ions. The configuration of the noble gas from the period in which the element is found is usually adopted as an energetically more favorable state.

In a methane molecule (CH_4) all the atoms have a noble-gas configuration. Carbon has in the p-orbitals two valence electrons and four free spaces for further electrons which are made available by the four hydrogen atoms. Carbon thus achieves the noble-gas configuration of neon. Because the two atoms share the common bonding-electron pair, the hydrogen atoms, which only have in each case one valence electron in the s-orbital, also attain the noble-gas configuration of helium through the common electron pairs with the carbon. In the case of the main-group elements of the 2nd period, the filling of empty orbitals with valence electrons of other atoms until the noble-gas configuration is attained is called the octet rule, because then a total of eight electrons, two in s- and six in p-orbitals, are present.

An atom can enter into several covalent bonds with a neighboring atom, such as for example carbon double bonds in ethylene ($H_2C=CH_2$) or triple bonds in acetylene ($H-C\equiv C-H$).

In the octet rule existing free electron pairs which have not bonding function are also taken into account. The three-dimensional structure of molecules is determined by the total number of atoms, by the type and number of bonding-electron pairs, and by the available free electron pairs. For example, methane (CH_4), ammonia (NH_3) and water (H_2O) have a tetrahedral basic structure. However, only the methane molecule represents an ideal tetrahedron with a bond angle of 109.5° (Figure 4). Free electron pairs require a greater amount of space then bonding-electron pairs. The available free electron pair in ammonia reduces the bond angle to 107.5°. The water molecule has two available free electron pairs, further reducing the bond angle to 104.5°.

Figure 4: Influence of free electron pairs on the structure of molecules
a) Methane (CH_4),
b) Ammonia (NH_3),
c) Water (H_2O).



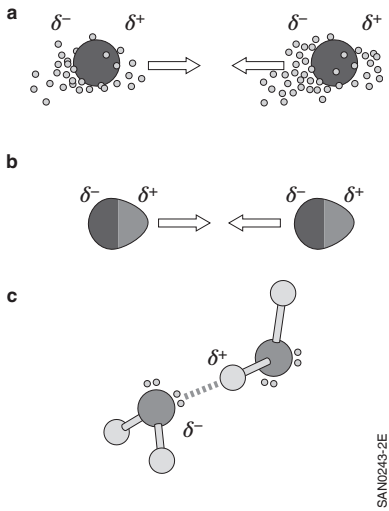
In the simplified model presented here covalent atomic bonds are put down to the formation of pairs of valence electrons from atomic orbitals. However, solely examining atomic orbitals only touches the surface of the complex conditions in a molecule. All the atomic nuclei and electrons in a molecule influence each other, and the bonding conditions are therefore better described by molecular orbitals. A quantum-mechanical examination of molecular orbitals is required to describe double and triple bonds and delocalized bonding systems in aromatic compounds; it is perfectly sufficient in this case to confine oneself to the valence electrons.

Metallic bonds
In metals the atoms are arranged in three-dimensional lattices where the valence electrons of the atoms move freely and are therefore often referred to as electron gas. Metals are good conductors of electricity and heat, which can be put down to the free valence electrons. The lattice vibrations of the atoms also control to the thermal conductivity; here the atoms vibrate freely in their positions in the metal lattice and can easily transmit heat in this way. Here too, a quantum-mechanical examination of molecular orbitals is required to better understand the greater complexities of the electric conductivity of metals. It is possible to derive from this that the electrons reside in energy bands which are formed from molecular orbitals with the smallest energy differences. Between the energy bands are zones which cannot be occupied by electrons.

Interactions between molecules
Van der Waals forces
Partial charges which are created by three-dimensional fluctuations of the positive and negative charge concentrations in the molecule induce electric dipoles (Figure 5a). Interactions that are created between the molecules polarized in this way are called van der Waals forces. The bigger the molecule, the more it can be polarized and the stronger the intermolecular forces.

Dipole-dipole interactions
In molecules which are composed of different atoms the charge concentration is continuously displaced (Figure 5b). Depending on the size and the charge of the atomic nuclei and inner electrons involved, the valence-electron pair is subjected to different degrees of attraction. The property of atoms to attract the bonding-electron pair is known as electronegativity. The more strongly attracting, more electronegative atom receives a partially negative charge. Correspondingly, the more weakly attracting atom receives a partially positive charge. The permanent dipoles brought about by a difference in the electronegativity exhibit intermolecular forces that are much stronger than van der Waals forces. These phenomena are called dipole-dipole interactions.

Figure 5: Types of interactions between molecules
a) Van der Waals forces,
b) Dipole-dipole interaction,
c) Hydrogen bond.
δ^+ Positive partial charge,
δ^- Negative partial charge.
Arrows indicate the force action.



The dipole-dipole interactions are also influenced by the three-dimensional structure of the molecule. Although there are differences in the electronegativity of carbon and oxygen, carbon dioxide (CO_2) for example is not a dipole on account of its linear structure. Water (H_2O) with the same atomic ratio 2:1, however, is a dipole, because the central oxygen atom has two free electron pairs, which result in an angled structure.

Hydrogen bonds
The angled structure of the water molecule facilitates additional dipole-dipole interactions (Figure 5c). The partially positively charged hydrogen atoms can interact with the free electron pairs of neighboring molecules. In liquid water the additional dipole-dipole interactions provide a particular short-range structure, which is responsible for water having the highest density at 4 °C. This enables fish to survive in winter because water with the highest density collects at the bottom of frozen lakes.

Molecular spectroscopy
Molecules, like atoms, can absorb energy and assume defined energy states. Valence electrons can be excited with ultraviolet or visible light (UV/VIS spectroscopy) and vibrations and rotations with infrared light (IR spectroscopy) respectively [6]. These procedures are used to explain structures because energy absorption allows conclusions to be drawn about molecular structure. In Raman spectroscopy the inelastic scattering of monochromatic light of molecules is analyzed. The frequency-shifted portions for example resulting from excitation of vibrations provide structural information.

Mass spectroscopy on the other hand is not based on the excitation of energy states. In this procedure molecules are ionized and the ions obtained are separated according to their mass-to-charge ratio and identified.

Substances

Substance term in chemistry
The chemical properties of a body are determined by its material and not by its size and shape. The material of a body is therefore also called substance. Even where the body has a very fine distribution the chemical properties remain the same. However, a large surface can increase the reactivity, as can be seen in nanoparticles.

Homogeneous and heterogeneous substances
Substances with a uniform structure are termed single-phase or homogeneous. A substance that is made up of two or more parts which cannot be mixed with each other is termed heterogeneous. An example of a homogeneous solid substance is elementary sulfur. Typical of a heterogeneous solid mixture is granite, which is made up of quartz, feldspar and mica.

Dispersion
Heterogeneous mixtures are always dispersions. Dispersions consist of at least two different substances which under the prevailing conditions do not or barely dissolve in each other and do not chemically react with each other. Depending on the phases, a distinction is made between suspensions (liquid and solid), emulsions (liquid and liquid) and aerosols (gas and solid or gas and liquid).

Suspension
If, for example, clay is added to pure water – a homogeneous liquid – this creates a heterogeneous mixture of a liquid and a solid which is termed a suspension.

Emulsion
Heterogeneous mixtures of two liquids, e.g. water and oil, are called emulsions.

Aerosol
A heterogeneous mixture of solid or liquid floating particles and a gas is called an aerosol. An example of an aerosol of gases and solids is exhaust gas with soot particles. Exhaust gas with white smoke, which is produced when water and sulfuric acid condense during the starting operation in the still cold exhaust gas, is on the other hand an aerosol of gases and liquids.

Colloid
The term colloid is used for particles or droplets ranging in size between 1 nm and 1 µm, irrespective of whether the heterogeneous mixtures are suspensions, emulsions or aerosols.

States of aggregation
The three classic states of aggregation (phases) are solid, liquid and gaseous – depending on whether the particles are in a fixed position in a solid body, can move in a liquid while maintaining a short-range order, or are far apart from each other in a gas. Plasma is a non-classic state of aggregation and consists of free electrons and ionized atoms.

The states of aggregation of substances are pressure- and temperature-dependent and are described in the state or phase diagrams. State diagrams, in which pressure is plotted against temperature, explain the conditions under which a solid is present, if necessary in different forms, the so-called modifications. Modifications refer to the phenomenon when a substance in the solid state occurs in different structural forms. Also following from the state diagrams are the ranges in which a liquid or a gaseous phase is present.

State diagram for carbon
The state diagram for carbon shows the solid modifications of graphite and diamond (Figure 6), where in ranges of these phases the other modification exists in parallel as a metastable form. Metastable means that the transformation of a modification is inhibited despite its higher energy content for small changes of state. Graphite is the stable modification at room temperature and atmospheric pressure; diamonds exist however as a metastable structure variant because the transformation into the more stable graphite form is strongly curbed by a high activation energy.

State diagram for water

The three areas in the state diagram for water show the existence ranges of ice, liquid water and vapor (Figure 7). Within these areas only the respective state of aggregation prevails. The curves between the areas describe the equilibrium between the liquid and gaseous phases (vaporization curve), the solid and liquid phases (fusion curve), and the solid and gaseous phases (sublimation curve). Each point on the curves corresponds to a state of equilibrium between the adjacent phases. At the point where the three curves intersect, the triple point TP (0.01 °C; 6.1 mbar), all three phases of water at equilibrium with one another. If the temperature or pressure is changed, only one state of aggregation still exists. If temperature and pressure are simultaneously changed in such a way that the new state corresponds to a point on one of the equilibrium curves, two phases exist next to each other.

An increase in temperature always results in a higher proportion of gaseous water, either through greater sublimation or through a higher vapor pressure. Conversely, a higher ambient pressure results in more condensation or resublimation of water vapor. The vapor-pressure curve demonstrates that the boiling point of water is dependent on external pressure. At low air pressure or in a vacuum water boils at temperatures below 100 °C.

What is conspicuous is the anomaly that can be observed at the phase transition from solid to liquid. An increase in both temperature and pressure causes ice to liquify. The transformation of ice into liquid water is furthered by pressure, because water assumes a smaller volume than ice.

A special thermodynamic state is the critical point CP (374 °C; 220.5 bar), at which the densities of the gaseous and liquid phases become indistinguishable. At the end of the vaporization curve the two states of aggregation – fluid and gaseous – pass into a new state, the supercritical phase. Water in the supercritical state is a liquid which has a lower density than liquid water below the critical point.

Figure 6: State diagram for carbon

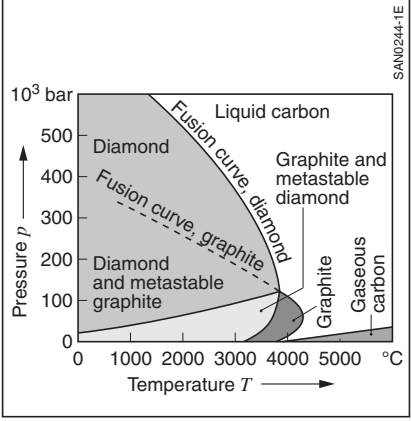
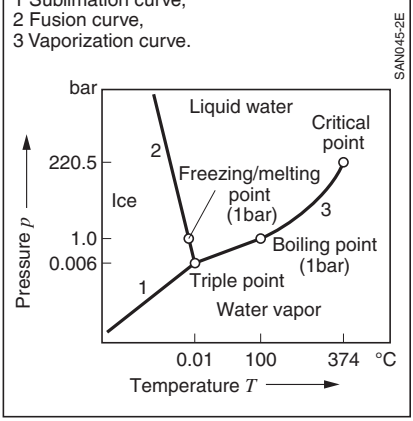


Figure 7: State diagram for water

1 Sublimation curve,
2 Fusion curve,
3 Vaporization curve.



Substance concentrations

To be able to describe the quantitative proportion of the substances involved in a chemical reaction, it is necessary to have quantitative details which refer to the number of particles involved in the conversion.

Owing to the different structure of atomic nuclei all element atoms and naturally also all the resulting molecular compounds differ in mass. It would not be very practicable to reckon with the minute masses of individual atoms or molecules directly. Because it is also helpful for chemical reactions always to consider the same number of particles, the term amount of substance was introduced with the unit "mole".

Amount of substance
An amount of substance n of 1 mole always contains the same number of particles and was therefore previously also often referred to as mole number. Irrespective of the chemical element, molecule or type of substance, 1 mole contains approximately $6 \cdot 10^{23}$ particles. This particle number can be experimentally determined and is also known as "Avogadro's number".

Molar mass
1 mole of a substance therefore always consists of around $6 \cdot 10^{23}$ particles. For element atoms the mass of these particles (molar mass M) therefore corresponds to their relative atomic mass, e.g. 1 mole of carbon has a mass of 12.011 g.

For chemical compounds the molar mass is calculated from the molar masses of the contained elements. Carbon dioxide (CO_2) has a molar mass of 44.01 g/mole (12.011 g/mole for carbon and $2 \cdot 15.9994$ g/mole for oxygen).

The molar mass of a substance – the mass of 1 mole of particles – therefore corresponds to a macroscopically manipulable amount of substance. The molar mass M is the quotient of the mass m and the amount of substance n of a substance (unit g/mol):

$$M = \frac{m}{n}.$$

In chemistry the term weight is also used instead of mass. Mass denotes the matter contained, weight the force acting as a result of the gravitational field on this matter. Since the force of gravity on Earth is approximately always the same, often no distinction is made between mass and weight. That is why the term "molar weight" is also widely used.

Molar volume
Since 1 mole of a substance always consists of the same number of particles, the volume (molar volume V_M) that these particles assume is always the same – provided the particles do not influence each other. This limiting case is only given for an ideal gas: 1 mole of an ideal gas assumes under normal pressure (p = 1013.25 mbar) at T = 273.15 K (0 °C) a volume of 22.414 l and at T = 298.15 K (25 °C) a volume of 24.789 l. With regard to the standard conditions, weight- and volume-specific concentration details about the molar volume can be approximately converted into each other for all gases.

Mole percent
Some concentration details are given in mole percent. A distinction is made here between the amount-of-substance- and volume-related definitions. The amount-of-substance-related concentration mole percent x with the unit % (n/n) is obtained by multiplying the mole fraction x_i – i.e. the amount of substance n_i of the constituent i referred to the sum of the amounts of substance of all the constituents of the substance mixture – by 100 %:

$$x = \frac{n_i}{\sum_{j=1}^{n} n_j} \cdot 100.$$

In the case of ideal gases the same particle number always assumes the same volume so that mole percent and percent by volume are identical.

Reactions of substances

Chemical thermodynamics
In chemical reactions starting substances are converted into reaction products with different properties. Chemical thermodynamics describes the substance conversion and the accompanying change in the internal energy ΔU, i.e. in the end whether and under what conditions reactions can take place [7]. The reaction path is determined by the amount of absorbed or output energy.

Heat of reaction
Most reactions are determine in open vessels, i.e. at a constant air pressure. At a constant pressure the change in the internal energy in a chemical reaction is made up of the two components heat of reaction Q_P and work W:

$$\Delta U = Q_P + W.$$

Mechanical work is done for example in an exothermic reaction in which a gas is produced which expands against atmospheric pressure or presses against a membrane or a moving punch.

Reactions can also be effected under a constant volume. Because then the energy content can only be changed by the heat of reaction Q_V and not by work W, the change in the internal energy ΔU corresponds to the heat of reaction Q_V. The following applies:

$$\Delta U = Q_V .$$

The heat of reaction heat at constant volume Q_V is therefore always greater than Q_P at constant pressure.

Enthalpy of reaction
The proportion of the heat of reaction Q_P or Q_V can also be described as the difference in heat contents of reaction and starting products and is called the enthalpy of reaction ΔH_R. Where heat is absorbed, the reaction is endothermic ($\Delta H_R > 0$). Chemical reactions in which heat of reaction heat is released ($\Delta H_R < 0$) are exothermic.

The relationship

$$\Delta U^0 = \Delta H_R^0 + W^0$$

is known as the first law of thermodynamics. As the change in the internal energy ΔU, the enthalpy ΔH_R and the work done W are dependent on the pressure and temperature, the standard conditions at 25 °C and 1,013 mbar are referenced and this is indicated at the symbols with a superscript zero (e.g. for ΔU^0).

Enthalpy of activation
Even exothermic reactions often have an initial energy demand despite the released enthalpy of reaction ΔH_R. This enthalpy of activation ΔH_A must be expended in order for example to break bonds and establish an activated complex of reacting agents before more energy is released by the creation of new bonds during the formation of the reaction product than was required to break the bonds in the starting products.

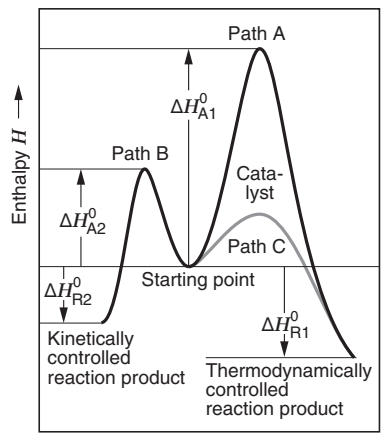
Thermodynamically and kinetically controlled reaction progression
The thermodynamically stable, i.e. energetically more favorable reaction product is reproduced in a reaction when the enthalpy of activation required for substance conversion is applied, e.g. through the supply of heat from the outside (Figure 8, path A). However, if the enthalpy of activation ΔH_{A1}^0 which is required to create the thermodynamically more stable product is not available, another chemical reaction may nevertheless under certain circumstances occur, namely the creation of the product kinetically controlled by the reaction rate. To create it an enthalpy of activation ΔH_{A2}^0 is likewise required (path B), but this is lower in terms of absolute amount. This product created by kinetic control of the reaction then has a higher internal energy. By choosing the reaction conditions, i.e. for example via the quantity of heat made available to the reaction, it is possible to influence whether the thermodynamically or kinetically controlled product tends to be created.

Catalyst decrease the enthalpy of activation of a reaction and thereby increase the reaction rate; but they cannot shift the thermodynamic equilibrium (path C).

Entropy

The enthalpy of reaction ΔH_R alone is not the only factor that determines the direction of a reaction. The distribution of the energy content also plays a role. The evaporation of a liquid takes place despite the required heat of evaporation $\Delta H_V > 0$ because the constraints for the individual molecules decrease as a result. The kinetic energy distributed to the individual molecules passes over from a more ordered, less probable distribution to a state of lesser order but greater probability. The change in the state of order is described by way of the characteristic entropy S [7]. Entropy increases in chemical reactions when more molecules are created than abreact, when the temperature rises, or when particles change to a less ordered state of aggregation (e.g. from solid to liquid or from liquid to gaseous). The increase in entropy ΔS is greater, the higher the heat input and the lower the temperature T at which the heat is transferred.

Figure 8: Enthalpy changes in thermodynamically and kinetically controlled reactions



Reaction kinetics

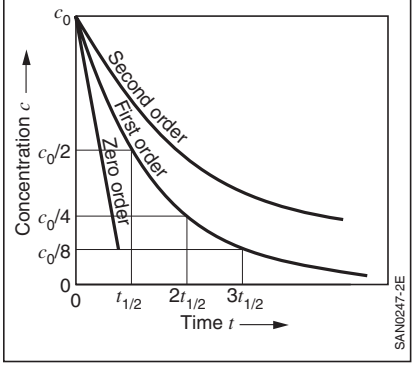
Whereas thermodynamics describes the conversion of substance and energy, reaction kinetics is concerned with the velocity or rate of reactions [7]. The reaction rate of a substances involved in the reaction is specified as a change in concentration per time interval (Figure 9). The order of a reaction is derived from the number of starting substances whose concentration changes during the reaction.

Zero-order reaction

A zero-order reaction exists for example when the decay of a gas on a platinum surface is heterogeneously catalyzed. The concentration of the gas adsorbed on the catalyst does not change during the reaction such that the reaction rate always remains the same, regardless of the reaction time.

Figure 9: Decrease in the concentration of a starting substance in reactions of different orders (assuming equal starting rates)

$t_{1/2}$ Half-life.



First-order reaction
In a first-order reaction the reaction rate depends only on the concentration of the or a starting product. Many decay processes such as radioactive decay for example follow a first-order reaction which is characterized by the fact that the time in which the concentration of the starting product decreases by half (half-life) is always the same. The half-life is therefore not dependent on the starting concentration.

Second-order reaction
A reaction between two molecules is called a second-order reaction when the reaction rate is determined by the concentration of the two starting substances. If one of the reactands is present in a very large surplus, as is the case for example in hydrolysis reactions in which water is simultaneously the solvent and the reacting agent, the reaction rate equates to that of a first-order reaction. This is called a pseudo-first-order reaction.

With equal starting concentration and starting rate a second-order reaction always runs more slowly than a first-order reaction. The rate of the two reaction types is dependent on the concentration of the starting substances. In a second-order reaction, however, it is necessary to factor in that for a substance transformation – unlike in a first-order reaction – a collision of molecules of the two reacting agents is required and also not every collision actually results in a transformation.

Third-order reaction
The probability of a trimolecular reaction, i.e. the simultaneous collision of three particles, is statistically low. For this reason, third-order reactions occur only very rarely.

Chemical equilibrium
Most chemical reactions are reversible, i.e. an equilibrium is obtained between the starting substances and the reaction products [8]. The position of equilibrium can naturally shift markedly from one side to the other and is influenced by a change in the concentration of a reacting agent, by removal of a reaction product, and by the choice of temperature.

In the reaction of two gaseous substances the conversion can be completed by increasing the pressure when the total number of molecules decreases due to the substance transformation. Low pressure would result in a worse conversion.

Similarly to pressure changes in gas reactions, changes in concentration also result in liquids. Dilution supports reactions which are accompanied by an increase in the number of particles. Catalysts on the other hand – as already described – cannot change the position of equilibrium, but can only influence the rate of its onset.

Law of mass action
Chemical equilibrium is described by the law of mass action [8]. The equilibrium constant K is obtained from the product of the concentrations c of substances C and D, which take part in the back reaction, divided by the product of the concentrations c of substances A and B, which are involved in the direct reaction. The law of mass action applies equally to dissolving processes, chemical reactions and changes of state:

$$A + B \rightleftharpoons C + D,$$

$$K = \frac{c(C) \cdot c(D)}{c(A) \cdot c(B)}.$$

When a salt is dissolved, this creates positively and negatively charged ions, i.e. cations and anions, which are in solubility equilibrium with the undissolved sediment. In a saturated solution the number of ions which dissolve equals the number of ions which precipitate again. x describes the number of cations A and y the number of anions B in the salt A_xB_y. The dissolving process sees the creation of x dissolved cations A, whose positive charge is determined by the number y of anions B. This is accompanied by the creation of y dissolved anions B with a negative charge, which is dependent on the number x of cations A.

$$A_xB_y \rightleftharpoons xA^{y+} + yB^{x-},$$

$$K = \frac{c(A^{y+})^x \cdot c(B^{x-})^y}{c(A_xB_y)}.$$

However, the law of mass action applies in this form only to the limiting case of an ideal solution without interaction between the dissolved particles. This is approximately the case with highly diluted solutions as are created when hardly soluble salts are dissolved.

Generally speaking, the law of mass action can only be applied to solutions when the concentrations c of ions are corrected by a factor f downwards. This factor, which takes into account the mutual influence of the dissolved particles, is concentration-dependent and is known as the activity coefficient f. Thus: $f \leq 1$. The activity coefficient takes into account the tendency of ions to associate in the solution due to the difference in charge. The concentration c corrected by the activity coefficient f is called the activity a. Thus:

$$a = f \cdot c,$$

$$K = \frac{a(A^{y+})^x \cdot a(B^{x-})^y}{a(A_xB_y)}.$$

No concentration can be specified for the undissolved portion. This portion is considered to be constant and set at one. In this simplified form the law of mass action is called the solubility product K_L:

$$K_L = a(A^{y+})^x \cdot a(B^{x-})^y.$$

The unit of the solubility product depends on the number of different particle types which are created during the dissolving process. The lower the solubility product, the harder the salt is to dissolve. For improved clarity, as a rule the decimal logarithm of the numerical value of K_L multiplied by –1 is specified instead of the solubility product and denoted as pK_L. The greater the pK_L value, the lesser the solubility.

The proportion of salt that dissolves decays into cations and anions. The process of splitting into separate ions of different charge is frequently called dissociation. The dissolving of salts creates many charge carriers which help to conduct electric current. In aqueous solutions chemists refer to strong electrolytes.

In daily laboratory practice calculations are made approximately with concentrations c instead of with activities a. The term "concentration" is used instead of "activity" in the following text, even if the activity is retained as the physical variable in the formulae.

Acids in water

Even acids (HA, A stands for acid) dissociate in water with the delivery of hydrogen ions H^+ to the water molecules and creation of hydroxonium ions H_3O^+ (protolysis). Here too the law of mass action can be applied [8].

$$HA + H_2O \rightleftharpoons H_3O^+ + A^-,$$

$$K = \frac{a(H_3O^+) \cdot a(A^-)}{a(HA) \cdot a(H_2O)},$$

$$K_a = K \cdot a(H_2O) = \frac{a(H_3O^+) \cdot a(A^-)}{a(HA)}.$$

The concentration of water with 55.3 mol/l remains virtually constant and is incorporated into the equilibrium constant, which is then called the acidity constant or K_a value and has the unit of a concentration in mol/l. When the K_a value is high, the equilibrium moves markedly to the right. The concentration of hydroxonium ions (H_3O^+) is high. The acid therefore has a high acidic strength.

The K_a value is as a rule converted into the pK_a value, whereby the K_a value is initially divided by the standard concentration of 1 mol/l. The decimal logarithm multiplied by –1 then produces the pK_a value. Strong acids have a low or even negative pK_a value.

Acidic strength of water
Even in chemically pure water the water molecules are present in equilibrium with hydroxonium ions (H_3O^+) and hydroxide ions (OH^-).

$$2\ H_2O \rightleftharpoons H_3O^+ + OH^-.$$

The ionic product K_w of water is always 10^{-14} mol^2/l^2. Autodissociation of water is therefore very low, in which hydroxonium ions (H_3O^+) and hydroxide ions (OH^-) are created in equally high concentration of 10^{-7} Mol/l.

Disregarding the water lost through dissociation, the acidic strength K_a of water ($a(H_2O)$ = 55.3 mol/l) is obtained at:

$$K = \frac{a(H_3O^+) \cdot a(OH^-)}{a(H_2O) \cdot a(H_2O)},$$

$$K_a = K \cdot a(H_2O) = \frac{a(H_3O^+) \cdot a(OH^-)}{a(H_2O)},$$

$$-\log K_a = -\log a(H_3O^+) - \log a(OH^-) + \log a\,(H_2O),$$

$$pK_a = 7 + 7 + \log 55.3 = 14 + 1.74 = 15.74.$$

Bases in water
Bases on the other hand, which dissolve in water, absorb hydrogen ions (H^+) [8]. When ammonia (NH_3) is dissolved in water, ammonia molecules adopt hydrogen ions (H^+) from the water molecules. Ammonium ions (NH_4^+) and hydroxide ions (OH^-) are created:

$$NH_3 + H_2O \rightleftharpoons NH_4^+ + OH^-.$$

According to the acidic constant K_a as the measure of acidic strength, it is also possible to describe the strength of bases by way of the base constant K_b. The basicity or base strength is derived from the law of mass action along similar lines to the acidic strength or more simply from the ionic product K_w of water via the relation

$$K_w = K_a \cdot K_b = 10^{-14}\ \text{mol}^2/l^2 \quad \text{or}$$

$$pK_w = pK_a + pK_b = 14.$$

Corresponding acid-base pairs
When an acid, e.g. hydrochloric acid (HCl), is dissociated, the hydrochloric acid represents the acid and the acid anion in equilibrium with the acid after hydrogen-ion delivery, in this example the chloride anion (Cl^-), the base. Together they are called a corresponding acid-base pair. Similarly, a base, e.g. ammonia (NH_3), in water, with its corresponding acid, the ammonium ion (NH_4^+), forms an acid-base pair.

The pK_a and pK_b values are experimentally determined for many acid-base pairs (Table 3). In the event of autodissociation of water, water acts as both an acid and a base. The property of water to act as both an acid and a base is known as amphoteric behavior. When water (pK_a = 15.74) is the acid, after hydrogen-ion delivery hydroxide ions (OH^-, pK_b = –1.74) are the corresponding base. When water (pK_b = 15.74) acts as a base, hydroxonium ions (H_3O^+, pK_a = –1.74) are the associated acid.

pH value
The pH value ("potentia hydrogenii") is defined as the negative decimal logarithm of the activity a^* of the hydroxonium ions (H_3O^+). To obtain a dimensionless quantity, the activity a of the H_3O^+ ions is divided before logarithmization by the standard concentration of 1 mol/l. The decimal logarithm then multiplied by –1 produces the pH value.

$$a^*(H_3O^+) = a(H_3O^+)\ [\text{mol}/l] \cdot \frac{1}{[\text{mol}/l]},$$

$$pH = -\log a^*(H_3O^+).$$

For simplification purposes the activity of the hydroxonium ions is often equated with the concentration.

The pH value is therefore a measure of the concentration of an acid; acid concentration (pH value) and acidic strength (pK_a value) do not necessary run parallel. A diluted hydrochloric-acid solution with a concentration of 10^{-4} mol/*l* can despite the high acidic strength ($pK_a = -6$) be more weakly acidic (pH = 4) than an acetic-acid solution ($pK_a = 4.75$) with a higher concentration of 10^{-1} mol/*l* (pH = 2.87) (Figure 10).

In the event of autodissociation of water, as already described above, hydroxonium ions (H_3O^+) and hydroxide ions (OH^-) in equal concentration of 10^{-7} mol/*l* are created. The pH value is therefore 7, pure water is neutral.

In the laboratory in actual fact pure water often reacts slightly acidly because CO_2 from the air dissolves with the formation of carbonic acid (H_2CO_3) in the water. Carbonic acid dissociates slightly to hydroxonium ions (H_3O^+) and hydrogencarbonate ions (HCO_3^-), pushing the pH value slightly in the acid direction.

The ratio of hydroxonium ions (H_3O^+) to hydroxide ions (OH^-) determines whether a solution is acidic, neutral or basic (Figure 10):

Acidic $a(H_3O^+) > a(OH^-)$: pH < 7.
Neutral $a(H_3O^+) = a(OH^-)$: pH = 7.
Basic $a(H_3O^+) < a(OH^-)$: pH > 7.

Figure 10: pH scale with examples of strong and weak acids and bases

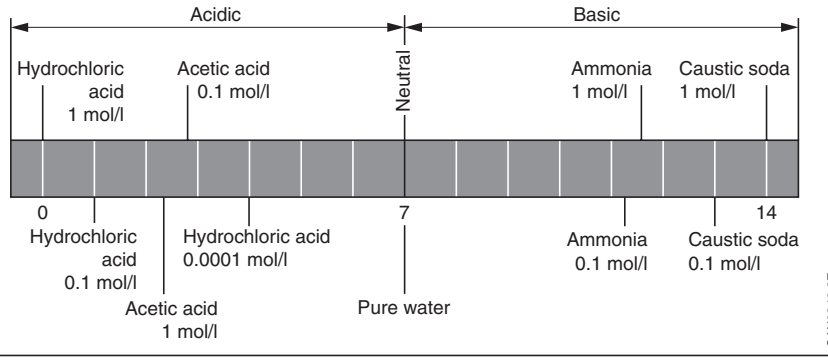


Table 3: pK_a and pK_b values of strong and weak acids

Name	Acidic strength pK_a	Corresponding acid-base pairs		Base strength pK_b
Hydrochloric acid	–6	HCl	Cl^-	20
Sulfuric acid	–3	H_2SO_4	HSO_4^-	17
Hydroxonium ion	–1.74	H_3O^+	H_2O	15.74
Nitric acid	–1.32	HNO_3	NO_3^-	15.32
Phosphoric acid	1.96	H_3PO_4	$H_2PO_4^-$	12.04
Hydrogen fluoride	3.14	HF	F^-	10.68
Acetic acid	4.75	CH_3COOH	CH_3COO^-	9.25
Carbonic acid	6.52	H_2CO_3	HCO_3^-	7.48
Ammonium ion	9.25	NH_4^+	NH_3	4.75
Hydrogen-carbonate ion	10.40	HCO_3^-	CO_3^{2-}	3.60
Hydrogen-phosphate ion	12.36	HPO_4^{2-}	PO_4^{3-}	1.64
Water	15.74	H_2O	OH^-	–1.74

pH value, strong acids

Strong acids like hydrogen chloride dissociate completely in water, during which they transfer positively charged hydrogen ions (H^+) to water molecules with the formation of hydroxonium ions (H_3O^+). 0.1 mol (3.65 g) hydrochloric acid in 1 liter of water produces 0.1 mol hydroxonium ions and 0.1 mol chloride anions (Cl^-). A 10^{-1} molar hydrochloric-acid solution therefore has a pH value of 1.

pH value, weak acids

In the case of weak acids, a certain amount of the acid remains undissociated in the water. But because the pH value of weak acids is only determined by the share of dissociated acid molecules, the calculation of the pH value is more complicated. This will be explained here using an acetic-acid solution by way of example:

In solutions of acetic acid in water less than 1% of all the acetic-acid molecules dissolved in water (CH_3COOH) are actually dissociated to hydroxonium ions (H_3O^+) and acetate ions (CH_3COO^-). In an aqueous acetic-acid solution, in contrast to an aqueous table-salt or hydrochloric-acid solution, the electrolyte is weak. The dissociation equilibrium can again be described using the law of mass action:

$$CH_3COOH + H_2O \rightleftharpoons H_3O^+ + CH_3COO^-,$$

$$K_a = \frac{a(H_3O^+) \cdot a(CH_3COO^-)}{a(CH_3COOH) \cdot a(H_2O)}.$$

The concentration of water can again be set at one on account of the large surplus.

Acetic acid has an acidic strength K_a of $1.8 \cdot 10^{-5}$ mol/l or put another way a pK_a value of 4.75. During dissociation as many hydroxonium ions (H_3O^+) as acetate ions (CH_3COO^-) are created. As the dissociation is minimal, the dissociated part compared with the total concentration of acetic acid CH_3COOH can be disregarded.

With $a(H_2O) = 1$ and $a(H_3O^+) = a(CH_3COO^-)$ it follows that:

$$K_a = \frac{a(H_3O^+)^2}{a(CH_3COOH)},$$

$$a(H_3O^+)^2 = K_a \cdot a(CH_3COOH),$$

$$a(H_3O^+) = \sqrt{K_a \cdot a(CH_3COOH)},$$

$$-\log a(H_3O^+) =$$

$$-\frac{1}{2}(\log K_a + \log a(CH_3COOH)) =$$

$$\frac{1}{2}(-\log K_a - \log a(CH_3COOH)).$$

Taking into account the definitions for the pH and pK_a values, the following is obtained:

$$pH = \frac{1}{2}(pK_a - \log a(CH_3COOH)).$$

An acetic-acid solution with a concentration of 0.1 mol/l therefore has a pH value of

$$pH = \frac{1}{2}(4.75 - \log 0.1)$$

$$= \frac{1}{2}(4.75 + 1) = 2.87.$$

pH values of other weak acids and bases can be similarly calculated.

Electrochemistry

When a salt is dissolved in water, the constituents of the salt are present in the water in the form of ions. These ions are charged in such a way that they move when an electric field is applied and thereby generate an electric current. This phenomenon is known as electrolytic conduction. In comparison, the current in an electric conductor such as copper or iron is transported by electrons. In addition to aqueous solutions, molten salts but also certain solids (such as zirconium oxide in the catalytic converter of a passenger car) can function as electrolytes ([9], [10] and [11]).

Negatively charged ions move towards the anode and are therefore called anions, while positively charged ions (cations) migrate towards the cathode. A chemical reaction takes place at these electrodes whereby the ions either absorb electrons from the cathode or discharge electrons to the anode. These reactions can only occur when the cathode and anode have an electrically conductive connection with each other in order to facilitate the exchange of electrons between the two.

If a battery is used as the voltage source, i.e., discharged, the electrons flow from the anode via the external electric circuit to the cathode. Thus, for the user the anode is the negative pole and the cathode is the positive pole.

Electrochemical series of metals

The intensity with which this ion reaction takes place is expressed by the electrochemical series of metals, which is set out in Table 4. The normal potential E^0, applicable to an ion concentration of 1 mol/*l*, is specified. This "normal" concentration is indicated by the superscript index 0. The voltages specified in the table refer to the normal hydrogen electrode, to which a potential of 0 V is therefore allocated. The following therefore applies:

$$E^0(2H^+ + 2e^- \leftrightarrow H_2) = 0\ \text{V}.$$

In Table 4, a positive sign with E^0 indicates an electron absorption (reduction), and a negative sign an electron discharge (oxidation). For example, lithium by preference discharges an electron ($E^0 < 0$) and is oxidized to become the singly charged lithium ion Li^+, while fluorine absorbs electrons ($E^0 > 0$) and is thus reduced.

Oxidation: $2\ Li \rightarrow 2\ Li^+ + 2e^-$,
$E_{Li} = +3.045\ \text{V}$.

Reduction: $F_2 + 2e^- \rightarrow 2\ F^-$,
$E_F = +2.87\ \text{V}$.

Balance equation: $2\ Li + F_2 \rightarrow 2\ Li^+ + 2\ F^-$,
$E = +5.915\text{V}$.

An electrochemical reaction therefore always comprises the two partial steps of oxidation and reduction. Because the lithium discharges the electrons, the electrons appear on the right side of the reaction equation compared with the indication in the electrochemical series of metals (Table 4). The sign preceding the voltage must therefore be reversed for this oxidation equation. The total voltage E of

Table 4: Electrochemical series of metals ([11]) with the associated ion reactions (normal potentials at 25 °C)

Half reactions	E^0 [V]
$Li^+ + e^- \leftrightarrow Li$	−3.045
$Na^+ + e^- \leftrightarrow Na$	−2.714
$Mg^{2+} + 2e^- \leftrightarrow Mg$	−2.363
$Al^{3+} + 3e^- \leftrightarrow Al$	−1.662
$2H_2O + 2e^- \leftrightarrow H_2 + 2OH^-$	−0.828
$Zn^{2+} + 2e^- \leftrightarrow Zn$	−0.763
$Cr^{3+} + 3e^- \leftrightarrow Cr$	−0.744
$Fe^{2+} + 2e^- \leftrightarrow Fe$	−0.440
$PbSO_4 + 2e^- \leftrightarrow Pb + SO_4^{2-}$	−0.356
$Ni^{2+} + 2e^- \leftrightarrow Ni$	−0.250
$Pb^{2+} + 2e^- \leftrightarrow Pb$	−0.13
$Sn^{2+} + 2e^- \leftrightarrow Sn$	−0.136
$2H^+ + 2e^- \leftrightarrow H_2$	0
$Cu^{2+} + 2e^- \leftrightarrow Cu$	+0.337
$Cu^+ + e^- \leftrightarrow Cu$	+0.521
$Fe^{3+} + e^- \leftrightarrow Fe^{2+}$	+0.771
$Ag^+ + e^- \leftrightarrow Ag$	+0.799
$Pt^{2+} + 2e^- \leftrightarrow Pt$	+1.118
$4H^+ + O_2 + 4e^- \leftrightarrow 2H_2O$	+1.229
$Cl_2 + 2e^- \leftrightarrow 2Cl^-$	+1.360
$Au^{3+} + 3e^- \leftrightarrow Au$	+1.498
$PbO_2 + 4H^+ + 2e^- \leftrightarrow Pb^{2+} + 2H_2O$	+1.685
$F_2 + 2e^- \leftrightarrow 2F^-$	+2.87

the redox reaction is thus added up from the individual values. The balance equation thus does not feature any electrons, as these are only exchanged between the reacting agents.

An electrochemical reaction can thus only take place when the total voltage is positive. An acid (i.e., H^+ ions) therefore cannot dissolve copper, silver, platinum, and gold. These are referred to as noble metals. In contrast, base metals such as sodium, elemental iron, nickel and lead are attacked by acids and the metals are dissolved as ions.

To enable an electrochemical reaction to take place, at least two different reacting agents are required. The resulting electrochemical total voltage E is dependent on the concentration of ions (for applications, see Lead battery, Lithium-ion cells).

Nernst equation

The Nernst equation reads:

$$E = E^0 + \frac{RT}{nF} \ln \frac{[\mathrm{Ox}]}{[\mathrm{Red}]} = E^0 + \frac{0.0592\ \mathrm{V}}{n} \log_{10} \frac{[\mathrm{Ox}]}{[\mathrm{Red}]} \quad (\text{at } 25\ ^\circ\mathrm{C}).$$

Here, [Ox] denotes the concentration of oxidized ions, [Red] the concentration of reduced ions, n the number of electrons in the reaction equation, R the gas constant, T the absolute temperature, and F the Faraday constant.

In this way, the potential can be calculated for each chemical partial reaction of the individual ion types. The total voltage of the electrolytic reactions is obtained from the sum total of all the potentials of the partial reactions. Over and above that E is dependent on the temperature.

Application

The oxygen concentration sensor (see λ sensors) measures the residual-oxygen content of the exhaust gas in the internal-combustion engine. To this end, a zirconium-oxide ceramic (ZrO) is used as a solid electrolyte, as this conducts oxygen ions at temperatures over 300 °C. Chemically inert platinum electrodes, which are applied as a porous thick layer, are used as electrodes. The voltage generated at the electrodes U_λ can be calculated with the Nernst equation. If here [Ox] and [Red] are replaced by the oxygen partial pressures $p(O_2)$ in the reference area (ambient air) and in the exhaust-gas area, the following relationship is obtained:

$$U_\lambda = \frac{RT}{4F} \ln \frac{p_\mathrm{R}(O_2)}{p_\mathrm{A}(O_2)}.$$

Here $p_\mathrm{A}(O_2)$ denotes the oxygen partial pressure in the exhaust gas, $p_\mathrm{R}(O_2)$ the oxygen partial pressure in the reference area, R the gas constant, T the absolute temperature, and F the Faraday constant.

References

[1] K.-H. Lautenschläger, W. Weber: Taschenbuch der Chemie. Edition Harri Deutsch, 21st Edition, 2013.
[2] M. Borlein: Kerntechnik – Grundlagen. Vogel Buchverlag, 2nd Edition, 2011.
[3] G. Schwedt: Analytische Chemie. Wiley-VCH Verlag, 2nd Edition, 2008.
[4] L. Spieß, G. Teichert, R. Schwarzer, H. Behnken, C. Genzel: Moderne Röntgenbeugung – Röntgendiffraktometrie für Materialwissenschaftler, Physiker und Chemiker. Springer Verlag, 3rd Edition, 2013.
[5] K. Schwister: Taschenbuch der Chemie. Carl Hanser Verlag, 4th Edition, 2010.
[6] M. Hesse, H. Meier, B. Zeeh: Spektroskopische Methoden in der Organischen Chemie. Thieme Verlag Stuttgart, 8th Edition, 2011.
[7] P.W. Atkins, L. Jones: Chemie – einfach alles. Wiley-VCH Verlag, 2nd Edition, 2006.
[8] R. Pfestorf: Chemie – Ein Lehrbuch für Fachhochschulen. Edition Harri Deutsch, 9th Edition, 2013.
[9] E. Fluck, R.C. Brasted: Allgemeine und Anorganische Chemie. 6th Ed., UTB No. 53, Quelle & Meyer, Heidelberg, 1987.
[10] C.E. Mortimer, U. Müller: Chemie. 12th Ed.,Thieme, Stuttgart, 2015.
[11] C.H. Hamann, W. Vielstrich: Elektrochemie. 4th Ed., Wiley-VCH, 2005.

Mathematics

Numbers

Quantities

Numbers are divided into natural numbers $\mathbb{N}$ = {0, 1, 2, 3, etc.}, whole numbers $\mathbb{Z}$ = {…, –3, –2, –1, 0, +1, +2, +3, …}, rational numbers $\mathbb{Q}$, real numbers $\mathbb{R}$ and complex numbers $\mathbb{C}$.

In addition to all whole numbers, rational numbers also include all fractions whose numerators and denominators are whole numbers. In addition to the rational numbers, real numbers also include all (infinitely many) numbers between fractions. Two examples of real numbers are the circular constant pi (π = 3.14159…) and the Eulerian number e (e = 2.718281…). Complex numbers are an extension of real numbers. These are explained in detail below (for more information, refer to [1], [2] and [3]).

Number systems

Physical and technical quantities are described by their numerical value and their unit. Numbers are usually represented in the decimal system, i.e. using base 10. Other common number systems are the binary and the hexadecimal systems, which are based on the numbers 2 and 16, respectively. While the decimal system features the digits 0, 1, 2 and onwards through 9, only the digits 0 and 1 exist in the binary system. The hexadecimal system uses, in addition to the digits 0 through 9, the letters A through F to represent the numbers 10 through 15 (for conversion, see Tables 1 and 2).

The binary and hexadecimal systems are used primarily in the field of information technology (IT), since a computer can only process the two states “power off” (“0”) and “power on” (“1”). These two states form the basis of the binary system. When eight binary digits are combined into one byte, it is then possible to represent the numbers 0 to 255, which corresponds to the hexadecimal digits 0 – FF.

Functions

The following section describes elementary mathematics functions. Their most important properties, such as their domain and value set, their behavior for very large and very small x values, their zero points, their derivations and the fundamental arithmetic operations with them are depicted.

These functions have been selected from a very large number of functions since they can be used to illustrate many technical processes, such as the geometric correlations between a connecting rod and moving mechanical parts (see Internal-combustion engines, Crankshaft gears), vibrations and oscillations in the vehicle (see Undercarriage, Basic principles of vibration characteristics), distance measurements of the vehicle relative to other cars and people (see Driver-assistance systems, Parking systems).

Table 1: Decimal and binary system

Decimal	Binary
0	0
1	1
2	10
3	11
4	100
8	1000
9	1001
15	1111
16	10000
32	100000
64	1000000
255	11111111

Table 2: Decimal and hexadecimal system

Decimal	Hexadecimal
0, 1 – 9	0, 1 – 9
10, 11 – 15	A, B – F
16	10
17	11
30	1E
31	1F
32	20
255	FF
4096	1000
65535	FFFF

Polynomial
A polynomial of the nth degree consists of $(n+1)$ summands with real (or complex) coefficients a_0, a_1,..., a_n and the associated monomials x^i:

$$f(x)=a_0+a_1x+a_2x^2+\ldots+a_nx^n,$$

where $a_i \in \mathbb{R}$ (or $a_i \in \mathbb{C}$)
and the domain range $D_f = \mathbb{R}$ (or $D_f = \mathbb{C}$)
and the value set $W_f = \mathbb{R}$.

A polynomial of the nth degree may have up to n zero points and $n-1$ local extrema.

Straight line
A straight line is a polynomial of the first degree ($n=1$):

$f(x)=a_0+a_1x$ or
$y=mx+t$ (with slope m and axis intercept t).

The zero point lies at $x_0=-\frac{t}{m}$ (for $m\neq 0$).

Parabola
A parabola is a polynomial of the second degree ($n=2$, quadratic function):

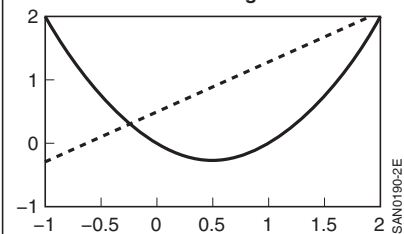
$f(x)=a_0+a_1x+a_2x^2$ or
$y=ax^2+bx+c.$

Zero points: $x_{1/2}=\frac{1}{2a}\left(-b\pm\sqrt{b^2-4ac}\right)$

Table 3: Polynomial

Definition range D_f, value set W_f
$f(x)=a_0+a_1x+a_2x^2+\ldots+a_nx^n$
$D_f=\mathbb{R}$, $W_f=\mathbb{R}$ (or subset).

Figure 1: Straight line (dashed) and parabola as polynomials of the first and second degree



A parabola may have zero, one or two zero points.

In Table 3 and Figure 1 shows by way of example a straight line (polynomial of the first degree) and a parabola (polynomial of the second degree).

Root function
The root function $f(x)=\sqrt{x}=x^{1/2}$ (Table 4 and Figure 2) is the inverse function of the quadratic function $f(x)=x^2$. It is needed for tasks such as solving quadratic equations (e.g. searching for the zero point of a second-degree polynomial, see Polynomial).

This also results in the root term, which is used in the calculating the resonant frequency and the damping ratio of an oscillation (see Oscillations).

Similarly, for each monomial $f(x)=x^n$, $n \in \mathbb{N}$, there is a strongly monotonically growing inverse function

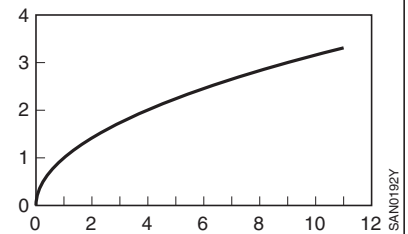
$$f^{-1}(x)=\sqrt[n]{x}=x^{1/n},$$

which is only defined for $x \geq 0$.

Table 4: Root function

Definition range D_f, value set W_f and behavior.	Properties.
$f(x)=\sqrt{x}=x^{1/2}$	$\sqrt{x}\cdot\sqrt{y}=\sqrt{xy}$
	$\sqrt{x^n}=(\sqrt{x})^n$, $n\in\mathbb{N}$
$D_f=\mathbb{R}_0^+$	
$W_f=\mathbb{R}_0^+$	
$x\to+\infty$: $f(x)\to+\infty$	

Figure 2: Curve of the root function



Absolute value and sign function

Each real number x can be broken down into its sign (+/–) and its absolute value $|x|$.

$x = \text{sgn}(x) \cdot |x|$.

For example, sgn(–3) = –1 and sgn(+4) = +1. The absolute value indicates the distance of the number x from a source of 0. The absolute value function (Table 5 and Figure 3) can be represented as

$$f(x) = |x| = \begin{cases} x & \text{for } x \geq 0 \\ -x & \text{for } x < 0 \end{cases}.$$

Exponential function

The exponential function is a very important function in mathematics, physics and technology, because it is the only function that is identical to its own derivative. This central property is used in solving linear differential equations such as the harmonics equation (see Oscillations).

Both the sine and cosine functions are related to the complex exponential function (see Complex numbers). This means the exponential function can be used to describe oscillations.

$f(t) = e^{-\gamma t + i\omega t} = e^{-\gamma t}$ (cos $\omega t + i$ sin ωt).

Here, the real negative exponent $-\gamma t$ represents the damping of the oscillation, while the complex exponent $i\omega t$ reflects the periodic quantity, which becomes clear from the representation of the sine and cosine functions (see Complex numbers, Linear differential equations).

Furthermore, laws of growth (e.g. interest and compound interest calculations) and laws of decay (e.g. radioactive decay) can be expressed with the exponential function. The charging and discharging processes of capacitors also follow an exponential curve (see Capacitor). The final compression pressure, final compression temperature and efficiency of a reciprocating-piston engine are exponentially related to the polytropes – more specifically, the adiabatic exponents (see Reciprocating-piston engine).

Table 6 and Figure 4 show the domain and value set, as well as the properties of the exponential function.

Table 5: Absolute value function

Definition range D_f, Value set W_f and behavior.
$f(x) = \text{sgn}(x)$ $D_f = \mathbb{R}$ $W_f = \mathbb{R}_0^+$ $x \to \pm\infty : \ f(x) \to +\infty$

Figure 3: Curve of the absolute value function

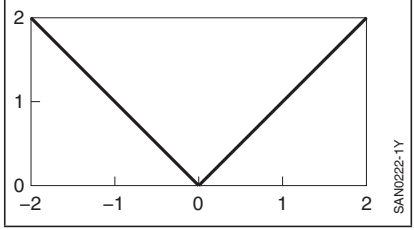
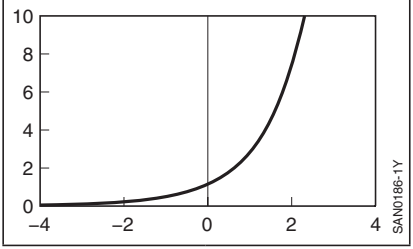


Table 6: Exponential function

Definition range D_f and behavior.	Properties.
$f(x) = e^x$; $D_f = \mathbb{R}$ $x \to -\infty : f(x) \to 0$ $x \to +\infty : f(x) \to +\infty$ $f(0) = e^0 = 1$	$e^a e^b = e^{a+b}$ $(e^a)^b = e^{ab}$ $\frac{d}{dx} e^x = e^x$
Euler's number $e = 2.71828...$	

Figure 4: Curve of the exponential function



Logarithm function

The logarithm function is the inverse function of the exponential function. This is needed in order to solve equations such as the following:

$$2^x = 8 \Rightarrow x = \log_2 8 \Rightarrow x = 3.$$

Table 7 and Figure 5 show the domain and value set, as well as the properties of the logarithm function.

Application examples

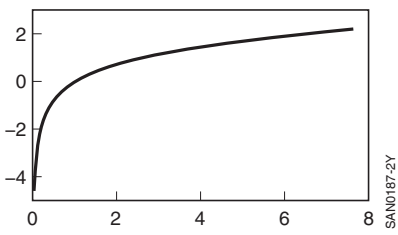
The Nernst equation provides one example of application of the logarithm function. This equation calculates the voltage value from the concentration of oxygen in the surroundings and the exhaust gas, which is necessary for λ regulation (see λ sensor).

In acoustics, decibel level (dB) is defined using the logarithm of the sound pressure (see Acoustics, Decibels). Similarly, logarithms are also used to determine the sound power level and sound intensity level.

Table 7: Logarithm function

Definition range D_f and behavior.	Properties.
$f(x) = \ln x$	$\ln a + \ln b = \ln ab$
$= \log_e x\ ;\ D_f = \mathbb{R}^+$	$\ln a - \ln b = \ln \frac{a}{b}$
$x \to 0:\ f(x) \to -\infty$	$c \ln a = \ln a^c$
$x \to +\infty:\ f(x) \to \infty$	$\frac{d}{dx} \ln x = \frac{1}{x}$
$f(1) = \ln 1 = 0$	

Figure 5: Curve of the logarithm function



Converting different logarithms

The logarithm of the number z for the base a can be converted to the base b as follows (a, $b > 0$):

$$\log_a z = \frac{\log_b z}{\log_b a}$$

This results in $\lg z = \log_{10} z$ for the common (decimal) logarithm and $\mathrm{lb}\, z = \log_2 z$ for the binary logarithm:

$$\log_a z = \frac{\ln z}{\ln a} = \frac{\mathrm{lb}\, z}{\mathrm{lb}\, a} = \frac{\lg z}{\lg a}.$$

Trigonometric functions

Angular measurement (radians)

In mathematics, angles are usually specified as the measure of an arc in radians and rarely in degrees. So angle $\varphi = 360°$ is equivalent to a arc of $x = 2\pi$. The circumference of a circle with a radius of 1 shares this value.

This yields the following conversions between angle φ in degrees and angle x in radians:

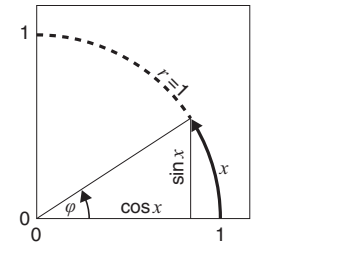
$$\frac{\varphi}{360°} = \frac{x}{2\pi},$$

$$\Rightarrow \varphi = \frac{180°}{\pi} x,$$

$$\Rightarrow x = \frac{\pi}{180°} \varphi.$$

The unit “rad” is often added to an arc measurement to make it clear that the information relates to an angle. Arc x of the angle φ is shown in Figure 6. Angle φ of the associated arc x is shown as a curved arrow.

Figure 6: Sine and cosine in a unit circle



Sine and cosine functions
In a right triangle, the sine of angle φ or arc x is equal to the ratio of the opposite side to the hypotenuse. The cosine is the ratio of the adjacent side to the hypotenuse.

In a right triangle with a hypotenuse with length $r = 1$ (unit circle), the adjacent side corresponds to the cosine and the opposite side corresponds to the sine (Figure 6).

Many periodic processes such as oscillations can be expressed through the sine or cosine function (Table 8 and Figure 7). If the variable x is replaced by

$$x = \frac{2\pi}{T}\,t \text{ and } \omega = \frac{2\pi}{T}$$

where the variable t usually stands for the time, then T is the period of the oscillation. The frequency f is the reciprocal of the period:

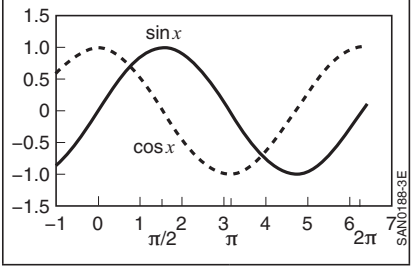
$$f = \frac{1}{T}$$

The angular frequency ω still contains the factor 2π; thus, it specifies the angle covered (in radians) per unit of time.

Table 8: Sine and cosine functions

Definition ranges D_f, D_g, Value sets W_f, W_g and behavior.	Properties.
$f(x) = \sin x\,;\quad D_f = \mathbb{R}$	$\sin(x + 2\pi) = \sin x$
$g(x) = \cos x\,;\, D_g = \mathbb{R}$	$\cos(x + 2\pi) = \cos x$
$W_f = W_g = [-1; +1]$	$\sin(x \pm \frac{\pi}{2}) = \pm\cos x$
Period length: 2π	$\cos(x \pm \frac{\pi}{2}) = \mp\sin x$
	$\sin^2 x + \cos^2 x = 1$

Figure 7: Curves of the sine and cosine functions



When a car drives on a street or a country road, it is constantly subjected to large and small bumps and shocks. Therefore, the undercarriage must be constructed so that it absorbs or compensates for the bumps in the road. In addition, the vibrational characteristics of the shock absorbers should be designed with precision (see Undercarriage, Basic principles of vibration characteristics).

Superpositions of oscillations are important in many applications, such as water waves, sound waves and the superposition of alternating electrical current (AC). In addition, the wave is usually formulated as a sine or cosine function

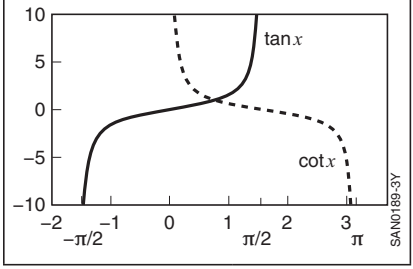
$f(t) = A\,\sin(\omega t + \varphi)$ or
$f(t) = A\,\cos(\omega t + \varphi)$

with (positive) amplitude A and phase shift φ.

Table 9: Tangent and cotangent functions

Definition ranges D_f, D_g, Value sets W_f, W_g, and behavior.	Properties.
$f(x) = \tan(x) = \frac{\sin(x)}{\cos(x)}$	$\tan(x + \pi) = \tan(x)$
$g(x) = \cot(x) = \frac{\cos(x)}{\sin(x)}$	$\cot(x + \pi) = \cot(x)$
	$\tan(x + \frac{\pi}{2}) = -\cot(x)$
	$\cot(x + \frac{\pi}{2}) = -\tan(x)$
$D_f = \mathbb{R} \setminus \{\frac{\pi}{2} + k\cdot\pi,\, k \in \mathbb{Z}\}$	$\cot(x) = \frac{1}{\tan(x)}$
$D_g = \mathbb{R} \setminus \{k\cdot\pi,\, k \in \mathbb{Z}\}$	
$W_f = W_g = \mathbb{R}$	
Period length: π	

Figure 8: Curves of tangent and cotangent functions



If oscillations with the same angular frequency are superimposed, they can be amplified (through constructive interference), weakened or even completely eliminated (through destructive interference). This depends on the phase relation and the amplitudes of the relevant oscillations.

For example, destructive interference is utilized as "anti-noise" at airports. This involves having speakers produce sounds with the same frequency as the turbine noise from airplanes. The anti-noise and noise are out of phase with one another, canceling each other out, reducing the noise for nearby residents.

Tangent and cotangent functions
In a right triangle, the tangent results from the ratio of the opposite side to the adjacent side. Therefore, the following relationship can be derived:

$$\tan x = \frac{\sin x}{\cos x}$$

The cotangent results from the ratio of the adjacent side to the opposite side. This means that

$$\cot x = \frac{\cos x}{\sin x}$$

Table 9 and Figure 8 show the properties and the curves of both of these functions. The most important properties of trigonometric functions are summarized in Table 10.

Arc functions
The arc functions are the inverse of trigonometric functions. For example, the following equation can be solved using the arcsine:

$\sin x = 0.5$ where $x \in [0; \pi/2]$

$\Rightarrow x = \arcsin(0.5) = \pi/6.$

Similarly, there are also an arccosine, arctangent and arccotangent.

In contrast to the sine, cosine, tangent and cotangent, which (with the exception of gaps) are defined for any $\mathbb{R}$, the arc functions may have a finite interval as a domain, in particular the value range of the original function. These properties and others are compiled in Table 11. The arc functions are not periodic.

Table 10: Properties of trigonometric functions, $k \in \mathbb{Z}$

	$\sin(x)$	$\cos(x)$	$\tan(x)$	$\cot(x)$
D_f	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}\setminus\{x \mid x=\frac{\pi}{2}+k\pi\}$	$\mathbb{R}\setminus\{x \mid x=k\pi\}$
W_f	$[-1; +1]$	$[-1; +1]$	$\mathbb{R}$	$\mathbb{R}$
Period	2π	2π	Π	Π
Symmetry	Odd	Even	Odd	Odd
Zero points	$x_0=k\pi$	$x_0=\frac{\pi}{2}+k\pi$	$x_0=k\pi$	$x_0=\frac{\pi}{2}+k\pi$
Maxima	$x_{max}=\frac{\pi}{2}+k\cdot 2\pi$	$x_{max}=k\cdot 2\pi$	–	–
Minima	$x_{min}=\frac{3}{2}+k\cdot 2\pi$	$x_{min}=\pi+k\cdot 2\pi$	–	–
Poles	–	–	$x_{Pole}=\frac{\pi}{2}+k\pi$	$x_{Pole}=k\pi$

Table 11: Properties of arc functions

	$\arcsin(x)$	$\arccos(x)$	$\arctan(x)$	$\operatorname{arccot}(x)$
D_f	$[-1;+1]$	$[-1;+1]$	$\mathbb{R}$	$\mathbb{R}$
W_f	$[-\frac{\pi}{2};\frac{\pi}{2}]$	$[0;\pi]$	$]-\frac{\pi}{2};\frac{\pi}{2}[$	$]0;\pi[$
Symmetry	Odd	Point symmetry for $P(0;\frac{\pi}{2})$	Odd	Point symmetry for $P(0;\frac{\pi}{2})$
Zero points	$x_0=0$	$x_0=1$	$x_0=0$	–
Monotony	Strictly monotonically increasing	Strictly monotonically decreasing	Strictly monotonically increasing	Strictly monotonically decreasing
Asymptotes	–	–	$y=\pm\frac{\pi}{2}$	$y=0;y=\pi$

Equations in the plane triangle

Figure 9 shows a triangle with any lengths of sides a, b and c and any angles α, β and γ. The relationship between these quantities is described by the following equations.

Angular sum:
$\alpha+\beta+\gamma=180°$.

Law of sines:
$a : b : c = \sin\alpha : \sin\beta : \sin\gamma$.

Law of cosines:
$c^2 = a^2 + b^2 - 2ab\cos\gamma$.

Pythagorean theorem:
$c^2 = a^2 + b^2$, where $\gamma = 90°$.

The relationships between angles in a triangle play an important role in the design of components. One example of this would be the gear design in a reciprocating piston engine, where the piston stroke and length of the connecting rod are related to one another by the angle of the crankshaft position and the leg angle of the connecting rod (see Internal-combustion engines, Gear design).

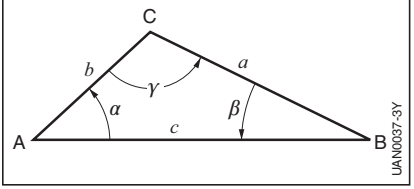
Complex numbers

In real numbers, a root can be extracted from any positive number; this is not possible with negative numbers. In order to bypass this restriction, real numbers are changed into complex numbers. Their central element is the imaginary unit i, where:

$i^2 = -1$.

A complex number z has a real component (Re) and an imaginary component (Im). The imaginary component is multiplied by the imaginary unit i. It is possible with the coordinates a and b to represent z in Cartesian form. Using a simple transformation, this can be converted into the polar coordinates r and φ (distance from the origin of the coordinates and the angle to the X-axis) (Figure 10).

Figure 9: Plane triangle
Vertices A, B and C.
Lengths of sides a, b and c;
Angles α, β and γ.



Complex number z in Cartesian form:
$z = a + ib$ where $i^2 = -1$.
$\mathrm{Re}\, z = a$, $\mathrm{Im}\, z = b$.

Complex number z in polar coordinates:
$z = re^{i\varphi}$

where $r = \sqrt{a^2 + b^2}$, $\tan\varphi = \frac{b}{a}$
$a = r\cos\varphi$, $b = r\sin\varphi$.

Complex exponential function (Euler's formula)
$e^{i\varphi} = \cos\varphi + i\sin\varphi$

Calculation rules for complex numbers
In the case of complex numbers, the same calculation rules that are used for real numbers apply, whereby addition is simpler to carry out in Cartesian notation and multiplication is simpler in polar notation.

Addition:
$z_1 + z_2 = (a_1 + a_2) + i(b_1 + b_2)$.

Multiplication:
$z_1 z_2 = (r_1 r_2)\, e^{i(\varphi_1 + \varphi_2)}$

In many cases, complex numbers aid in making complicated mathematical problems easier to solve. Frequently, only the real part or the imaginary part of the complex solution is necessary in finding the real solution.

For example, oscillation processes can be expressed using linear differential equations (see Differential equations), such as the oscillation and damping properties of shock absorbers or the current strength and voltage in an alternating-current circuit. A solution can be found quickly with the help of a complex exponential function. Finally, Euler's formula for exponential functions is used to convert a complex solution into a real solution.

Coordinate systems

Coordinate systems on a plane
In the previous segment, a complex number z is either represented by both of its coordinates x and y or by its absolute value $r = |z|$ and the angle φ in relation to the x-axis. This number can be graphically represented as a point on a plane (Figure 10).

Similarly, all the points on a plane can be expressed by their Cartesian coordinates (x and y) or by their polar coordinates (r and φ), and converted into one another.

Conversion of Cartesian coordinates to polar coordinates:

$r = \sqrt{x^2 + y^2}$, $x, y \in \mathbb{R}$,

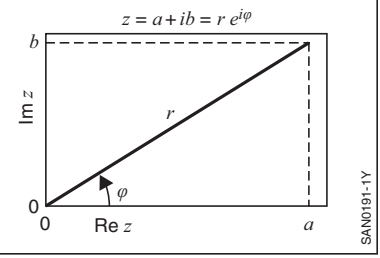
$\tan\varphi = \frac{y}{x}$.

Conversion of polar coordinates to Cartesian coordinates:
$x = r\cos\varphi$, $r \in \mathbb{R}_0^+$, $\varphi \in [0; 2\pi]$,
$y = r\sin\varphi$.

This conversion always yields a unique value, meaning that for each x-y pair, there is an exact r-φ pair. r is only = 0 at the coordinate origin (0, 0), but the angle φ is undefined. However, in practice this does not amount to a limitation.

Cartesian coordinates can be used for straight-line movements. Naturally, circular motion can also be represented us-

Figure 10: Complex number z
Diagram in Cartesian coordinates (a and b) and polar coordinates (r and φ).



ing Cartesian coordinates. However, the corresponding equations are significantly more complicated in comparison to a description in polar coordinates (see for example Fundamentals of vehicle engineering, Translational movement).

Coordinate systems in three dimensions

Each point in a three-dimensional space has three Cartesian coordinates (x,y,z). This representation is used frequently when the system to be described is at a right or oblique angle. However, if the problem relates to a rotationally symmetrical or a spherically symmetrical system, using cylindrical or spherical coordinates is recommended.

In Figure 11, the x, y and z coordinate axes are delineated. Here, the distance r of point P to the origin and the two angles θ and φ, which can be measured from the z or the x-axis. r, θ and φ are labeled as spherical coordinates or spatial polar coordinates.

With the exception of the origin (0, 0, 0), the coordinates for each point are clearly defined in both Cartesian and polar coordinates. Conversion between both coordinate system occurs as follows.

Conversion of Cartesian coordinates to polar coordinates:

$$r = \sqrt{x^2 + y^2 + z^2}\ ,\ x, y, z \in \mathbb{R},$$

$$\tan \varphi = \frac{y}{x},$$

$$\cos \Theta = \frac{z}{r} = \frac{z}{\sqrt{x^2+y^2+z^2}}\ .$$

Conversion of polar coordinates to Cartesian coordinates:

$$x = r\cos\varphi \sin\Theta,$$
$$r \in \mathbb{R}_0^+,\ \varphi \in [0; 2\pi],\ \Theta \in [0; \pi],$$
$$y = r\sin\varphi \sin\Theta,$$
$$z = r\cos\Theta\ .$$

The surface of the Earth is broken down cartographically into geographic coordinates. The geographic length is the angle φ in spherical coordinates. The geographic width is determined by an angle of −90° to +90° relative to the equator. This is equivalent to the angle θ in spherical coordinates, which is measured starting from the north pole and is between 0° and 180° (0 and π in radians).

Cylindrical coordinates

If an object or a system has rotational symmetry, cylindrical coordinates are utilized (Figure 12). The x and y coordinates, just like the polar coordinates on a plane, are converted to the radius ρ; the z component of the cylindrical coordinates

Figure 11: Spherical coordinates of a point

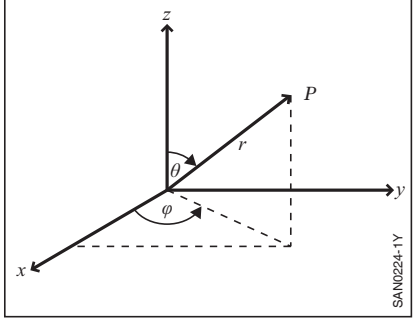
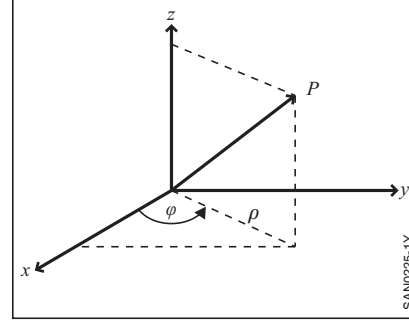


Figure 12: Cylindrical coordinates of a point



is identical to the z of the Cartesian coordinates.

Just like the spherical coordinates, any point can be clearly expressed in cylindrical coordinates, with the exception of the origin (0, 0, 0).

Conversion of Cartesian coordinates to cylindrical coordinates:

$$\rho = \sqrt{x^2 + y^2}, \quad x, y, z \in \mathbb{R},$$

$$\tan \varphi = \frac{y}{x},$$

$$z = z.$$

Conversion of cylindrical coordinates to Cartesian coordinates:

$$x = \rho \cos \varphi, \quad \rho \in \mathbb{R}_0^+, \varphi \in [0; 2\pi], z \in \mathbb{R},$$
$$y = \rho \sin \varphi,$$
$$z = z.$$

Just like the polar coordinates, cylindrical coordinates are calculated on a plane and they are expanded by the addition of the z-axis.

Cylindrical coordinates can be used in the design of rotationally symmetrical objects, such as tubes, engine cylinders, screws, nuts, and anti-friction and friction bearings.

Vectors

A distinction is made between scalar and vector variables for physical and technical concepts. Examples of scalars include mass, temperature and pressure. By comparison, vectors have a direction, such as velocity, power and electric field. Vectors, their most important arithmetic operations and calculation rules are shown below.

Representation of vectors
A vector in a three-dimensional space has the three components a_x, a_y and a_z for the x-, y and z direction. In mathematics, this is described as follows:

$$\vec{a} = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix}.$$

Calculation rules
When adding and multiplying vectors with a number, the same calculation rules for adding and multiplying numbers apply.

Addition
For adding two vectors with coordinates as follows

$$\vec{a} = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} \text{ and } \vec{c} = \begin{pmatrix} c_x \\ c_y \\ c_z \end{pmatrix}$$

results in:

$$\vec{a} + \vec{c} = \begin{pmatrix} a_x + c_x \\ a_y + c_y \\ a_z + c_z \end{pmatrix}$$

The following laws apply:
– Closure property: The sum and the difference of two vectors is also a vector.
– Commutative property:
$\vec{a} + \vec{c} = \vec{c} + \vec{a}$.
– Associative property:
$\vec{a} + (\vec{c} + \vec{n}) = (\vec{a} + \vec{c}) + \vec{n}$.
– For any two vectors $\vec{a}$ and $\vec{c}$ there is always one vector $\vec{z}$ so that:
$\vec{a} + \vec{z} = \vec{c}$, in other words $\vec{z} = \vec{c} - \vec{a}$.

Multiplication of a vector with a scalar
When multiplying vector a by scalar $\lambda \in \mathbb{R}$, the following is true:

$$\lambda \vec{a} = \begin{pmatrix} \lambda a_x \\ \lambda a_y \\ \lambda a_z \end{pmatrix}$$

For vector $\vec{c}$, the following is true:

$$\vec{c} = \lambda \vec{a} \Rightarrow |\vec{c}| = |\lambda| \cdot |\vec{a}|.$$

For vectors $\vec{a}$ and $\vec{c} \in \mathbb{R}^3$ and scalars λ and $\mu \in \mathbb{R}$, the following laws apply:
– Closure property: The product of a vector and a scalar is also a vector.
– Associative property: $(\lambda \cdot \mu) \cdot \vec{a} = \lambda \cdot (\mu \cdot \vec{a})$.
– Distributive law:
$(\lambda + \mu) \cdot \vec{a} = \lambda \cdot \vec{a} + \mu \cdot \vec{a}$,
$\lambda \cdot (\vec{a} + \vec{c}) = \lambda \cdot \vec{a} + \lambda \cdot \vec{c}$.
– Multiplying by the neutral element:
$1 \cdot \vec{a} = \vec{a}$.

Scalar product
When multiplying two vectors with one another, two different products are defined: the scalar product and the vector product.

The result of the scalar product is a number (scalar). For example, physical work is defined as a scalar product of the force vector $\vec{F}$ and the path vector $\vec{s}$ (Figure 13). Only the components of $\vec{F}$ that point in the direction of $\vec{s}$ contribute to the work. This corresponds to the offset projection depicted in Figure 13. Thus, the following is true for the absolute value of the projection

$$|\vec{F}_s| = |\vec{F}| \cos \varphi$$

The scalar product is defined as

$$\vec{a} \cdot \vec{c} = |\vec{a}| \cdot |\vec{c}| \cdot \cos \varphi$$
$$= a_x c_x + a_y c_y + a_z c_z.$$

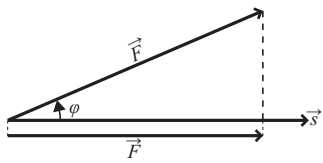
In order to clarify the product, the multiplication point "$\cdot$" is placed between both vectors.

The calculation rules for the scalar product are the same as the calculation rules for numbers.
– Commutative property:
$\vec{a} \cdot \vec{c} = \vec{c} \cdot \vec{a}$.
– Distributive law:
$\vec{a} \cdot (\vec{c} + \vec{n}) = \vec{a} \cdot \vec{c} + \vec{a} \cdot \vec{n}$.
– Multiplication with a scalar, associative property:
$\lambda (\vec{a} \cdot \vec{c}) = (\lambda \vec{a}) \cdot \vec{c} = \vec{a} \cdot (\lambda \vec{c})$.

For example, the scalar product is used to calculate the absolute value (the length) of a vector. In addition, the angles between the vectors can be determined using the same method. The scalar product in particular is used in order to find out if two vectors are perpendicular to each other. In this case, the scalar product is zero.

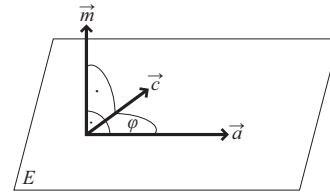
Cross product
A new vector results from the cross product (outer product) of two vectors; this new vector is perpendicular to both output vectors (Figure 14). Its length equals the area of a parallelogram spanning both initial vectors. The vectors $\vec{a}$, $\vec{c}$ and $\vec{m}$ form what is known as a right-handed system. According to the right-hand rule, the thumb points in the direction of $\vec{a}$, the index finger

Figure 13: Scalar product
The scalar product of both vectors $\vec{F}$ and $\vec{s}$ with the projection $\vec{F}_s$ of the vector $\vec{F}$ on $\vec{s}$. φ is the angle between both vectors $\vec{F}$ and $\vec{s}$.



SAN0226Y

Figure 14: Cross product
Both vectors $\vec{a}$ and $\vec{c}$ span the length of a plane E. The cross product (vector m) formed from $\vec{a}$ and $\vec{c}$ points perpendicular to both output vectors. φ is the angle between $\vec{a}$ and $\vec{c}$.



SAN0227Y

in the direction of $\vec{c}$ and the resulting vector $\vec{m}$ in the direction of the middle finger on the right hand.

In order to be able to distinguish the cross product from the scalar product, an "x" is placed between the vectors as a multiplication symbol.

The calculation rules for the cross product are summarized below:

$\vec{m} = \vec{a} \times \vec{c}$ where

$$|\vec{m}| = |\vec{a}| \cdot |\vec{c}| \cdot \sin\varphi = \begin{vmatrix} \vec{e}_x & \vec{e}_y & \vec{e}_z \\ a_x & a_y & a_z \\ c_x & c_y & c_z \end{vmatrix} = \begin{bmatrix} a_y c_z - a_z c_y \\ a_z c_x - a_x c_z \\ a_x c_y - a_y c_x \end{bmatrix}$$

$\vec{m} \perp \vec{a}$, $\vec{m} \perp \vec{c}$,

$\vec{a}$, $\vec{c}$, $\vec{m}$ are "right-handed"

with the unit vectors $\vec{e}_x$, $\vec{e}_y$ and $\vec{e}_z$ in x, y and z direction.

When applying the calculation rules, note that the leading sign is inverted if the order of the vectors is reversed (anticommutativity):

– Anticommutativity:
  $\vec{a} \times \vec{c} = -\vec{c} \times \vec{a}$.
– Distributive law:
  $\vec{a} \times (\vec{c} + \vec{n}) = \vec{a} \times \vec{c} + \vec{a} \times \vec{n}$.
– Multiplication with a scalar, associative property:
  $\lambda(\vec{a} \times \vec{c}) = (\lambda \vec{a}) \times \vec{c} = \vec{a} \times (\lambda \vec{c})$.

The cross product becomes zero when both vectors point in the same or opposite direction (i.e. when they are collinear). This lets you use the cross product to check if two vectors span one plane. In this case, the cross product does not equal zero.

Therefore, the cross product of a vector with itself is zero.

$\vec{a}, \vec{c} \neq \vec{0}$ where $\vec{a} \times \vec{c} = \vec{0} \Leftrightarrow \vec{a} \parallel \vec{c}$,

$\vec{a} \times \vec{a} = \vec{0}$.

Differential and integral calculus

Differentiation of functions

First derivative

The zero points, discontinuities and the behavior of a function can be determined at the margins of the domain using the function term $y = f(x)$. Finding out how steep a function is or where the function attains its maximum or minimum values is often of interest as well. This can be described using differentiation (the derivative) of a function.

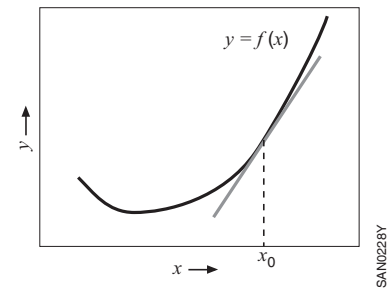
The slope (steepness) of a function $y = f(x)$ is given via the tangent of this function (Figure 15). If a function attains a maximum or a minimum value, then the tangent is horizontal at that point. In order to calculate the slope of a function $f(x)$, it must be derived (differentiated). The following two notation systems have been established for the first derivative of a function $f(x)$ to accomplish this:

$$\frac{d}{dx} f(x) = f'(x).$$

The derivatives of elementary functions are summarized in Table 12.

The derivative of compound functions can be calculated using the following calculation rules:

Figure 15: Derivative of a function



Sum rule:

$$\frac{d}{dx}(f(x) \pm g(x)) = f'(x) \pm g'(x).$$

Multiplication with a number λ:

$$\frac{d}{dx}(\lambda f(x)) = \lambda f'(x).$$

Product rule:

$$\frac{d}{dx}(f(x) \cdot g(x)) = f'(x) \cdot g(x) + f(x) \cdot g'(x).$$

Quotient rule:

$$\frac{d}{dx}\left|\frac{f(x)}{g(x)}\right| = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{[g(x)]^2}$$

Chain rule:

$$\frac{d}{dx}(F(g(x))) = F'(g(x)) \cdot g'(x).$$

Higher derivatives
If the first derivative is differentiated again, then this becomes the second derivative. Higher derivatives can be formed in a similar way.

2nd derivative:

$$\frac{d}{dx}f'(x) = f''(x) \text{ or } f''(x) = \frac{d^2}{dx^2}f(x)$$

3rd derivative:

$$\frac{d}{dx}f''(x) = f'''(x) \text{ or } f'''(x) = \frac{d^3}{dx^3}f(x).$$

nth derivative($n > 3$):

$$\frac{d}{dx}f^{(n-1)}(x) = f^{(n)}(x) \text{ or } f^{(n)} = \frac{d^n}{dx^n}f(x).$$

Extrema of functions
If the first and the second derivative of a function $f(x)$ are on hand, then these can be used to determine the maxima and minima (extrema) of $f(x)$. Both of the following conditions apply to an extremum at x_0:

– $f'(x_0) = 0$,
– $f''(x_0) < 0$ for a maximum,
  $f''(x_0) > 0$ for a minimum.

Inflection point
The curvature properties of the function can be determined by using the second derivative. The point x_w at which a function switches from left-curved ($f''(x) > 0$, convex curvature) to right-curved ($f''(x) < 0$, concave curvature) and vice versa is of particular interest. These points are known as inflection points. The following is true here:

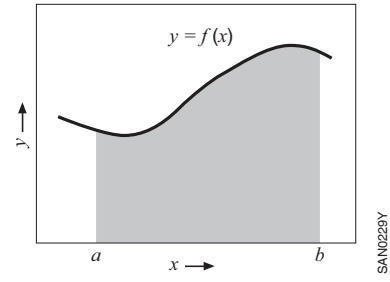
$$f''(x) = 0.$$

Integration of a function
A (continuous) function $f(x$ is depicted in Figure 16. Together with the X-axis in

Table 12: Elementary functions $f(x)$ and their derivatives $f'(x)$

$f(x)$	$f'(x)$
c = const	0
x^n	nx^{n-1}
$\sqrt{x} = x^{1/2}$	$\frac{1}{2}\frac{1}{\sqrt{x}} = \frac{1}{2}x^{-1/2}$
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$
$\tan(x)$	$\frac{1}{\cos^2(x)}$
$\cot(x)$	$-\frac{1}{\sin^2(x)}$
e^x	e^x
$a^x = e^{x \ln a}$	$(\ln a)\, a^x$
$\ln x$	$\frac{1}{x}$
$\log_a x = \frac{\ln x}{\ln a}$	$\frac{1}{\ln a}\frac{1}{x}$

Figure 16: Integral
Integration of the function $f(x)$ in the interval of a through b.



the interval [a; b], this encloses the area A marked in gray. Its dimensions can be calculated through integration.

The unknown area A is represented as an integral from a to b via the function $f(x)$ with the variable x:

$$A = \int_a^b f(x)\,dx.$$

Antiderivative

In order to find a solution to this problem, the antiderivative $F(x)$ is required for $f(x)$. Its property is expressed in the fundamental theorem of differential and integral calculus.

$$F(x) = \int_a^x f(t)\,dt$$

is an antiderivative for $f(x)$, if the following applies:

$$\frac{d}{dx}F(x) = f(x).$$

If $F(x)$ is an antiderivative for $f(x)$, then the unknown area can be calculated as

$$A = \int_a^b f(x)\,dx = F(b) - F(a).$$

Important antiderivatives are compiled in Table 13. For each function $f(x)$, there is just one antiderivative $F(x)$. This is uniquely defined with the exception of the integration constant C.

Table 13: Elementary functions and their antiderivatives (examples)

$f(x)$	$\int f(x)\,dx$
0	$c = \text{const}$
$x^n,\ c \neq -1$	$\frac{1}{n+1}x^{n+1} + C$
$\frac{1}{x}$	$\ln\lvert x\rvert + C$
$\sqrt{x} = x^{1/2}$	$\frac{2}{3}x^{3/2} + C$
$\sin x$	$-\cos x + C$
$\cos x$	$\sin x + C$
$\frac{1}{\cos^2 x}$	$\tan x + C$
$\frac{1}{\sin^2 x}$	$-\cot x + C$
e^x	$e^x + C$
$a^x = e^{x \ln a}$	$\frac{a^x}{\ln a} + C$

Calculation rules for integration

The following calculation rules apply to integration:

Sum rule:

$$\int_a^b f(x) + g(x)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx.$$

Multiplication with a number λ:

$$\int_a^b \lambda f(x)\,dx = \lambda \int_a^b f(x)\,dx.$$

Reversing the bounds of integration:

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx.$$

Lower parts of integration ranges in intervals ($a < c < b$):

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

Upper and lower bounds of integration are identical:

$$\int_a^a f(x)\,dx = 0.$$

Partial integration

In addition to searching for the antiderivative, there are two helpful methods for solving for integrals for partial integration and integration through substitution. Partial integration is the inverse of the product rule for the derivative.

$$\int_a^b f'(x)\cdot g(x)\,dx = f(x)\cdot g(x) - \int_a^b f(x)\cdot g'(x)\,dx.$$

The critical factor for this method is properly assigning both product terms $f'(x)$ and $g(x)$ within the integral.

Substitution

Substitution is the inverse of the chain rule. The art of substitution lies in finding the appropriate function that can be replaced (for more information, refer to [1], [2], [3]).

Linear differential equations

In many technical problems, a function is being sought that can describe the task but is not explicitly specified. Instead, information about its curvature properties (second derivative), its slope (first derivative) or a combination of the two is provided in the form of an equation.

An equation in which the function and its derivatives can be found is known as a differential equation. Equations known as common linear differential equations of the nth order with constant coefficients are covered below ([1], [2] and [3]):

$$y^{(n)}(x) + a_{n-1}y^{(n-1)}(x) + \ldots + a_0 y(x) = g(x).$$

In this equation, the nth derivative of $y(x)$ is the highest order derivative. Real numbers stand in front of individual derivatives $a_0, a_1, \ldots, a_{n-1}, a_n$, where $a_n = 1$ is selected. If the function $g(x)$ is always zero, the differential equation is homogeneous; otherwise, it is non-homogeneous. Here, x is the variable of the unknown function $y(x)$.

The differential equation above is linear, since $y(x), y'(x), \ldots y^{(n)}(x)$ are only accompanied by the coefficients and do not occur as exponents, such as $y(x)^2$, or as a nonlinear function, like $\sin(y'(x))$.

Using the formulation
$y(x) = Ae^{\lambda x}$, $A \in \mathbb{R}$, $\lambda \in \mathbb{R}$, a homogeneous differential equation is simplified to a characteristic polynomial, because the exponential term is canceled out:

$$\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_2\lambda^2 + a_1\lambda + a_0 = 0.$$

In doing so, solving the differential equation is reduced to searching for the zero points of the characteristic polynomial. If the n zero points $\lambda_1, \lambda_2, \ldots, \lambda_n$ are determined, then the homogeneous solution $y_h(x)$ of the differential equation is yielded as

$$y_h(x) = A_1 e^{\lambda_1 x} + A_2 e^{\lambda_2 x} + \ldots + A_n e^{\lambda_n x}.$$

The numbers $A_1, A_2, \ldots A_n$ are (any) real numbers and are labeled as integration constants. If initial or ancillary conditions are given for the function, then the values $A_1, A_2, \ldots, A_n$ can be calculated from these conditions.

After finding the homogeneous solution, a special solution $y_{ih}(x)$ for the non-homogeneous differential equation needs to be found; this then provides the complete solution.

$$y(x) = y_h(x) + y_{ih}(x).$$

Linear differential equations of the second order with constant coefficients are the basis of all oscillation processes (see Vibrations and oscillations, Vibration characteristics) and also the basis for the design of safety-related sensors like acceleration and vibration sensors (see Acceleration sensors). Even regulating and control engineering technology is based on these differential equations (see Control engineering).

In addition to common linear differential equations with constant coefficients, there are also linear differential equations in which the constant coefficients are replaced by functions, such as

$$\sin(x) \cdot y'(x) + \cos(x) \cdot y(x) = 0.$$

Furthermore, the two-dimensional oscillation equation of the function $f(x,y)$ with the location coordinates x and y and time t,

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = \frac{1}{c^2}\frac{\partial^2 f}{\partial t^2}, \; c \in \mathbb{R},$$

is an example of a partial differential equation. It is significantly more complicated to solve both of these types of differential equations than it is to solve the linear differential equations described above. For more information and a more detailed approach to these differential equations, the extensive literature from the bibliography should be consulted, including [1], [2] and [3], or the specialized literature regarding special types of differential equations.

Laplace transform

There are many control loops in a vehicle (see Control engineering); in the engine (e.g. knock control, λ control), in the A/C unit or in the undercarriage (e.g. yaw-rate control). These control loops are often represented by linear differential equations. One way to solve this differential equation is to use the exponential approach (see Differential equations).

Alternatively, the Laplace transform can also be utilized if initial values like $y(0)$, $y'(0)$ etc. are given in addition to the differential equation ([1], [2], [3]). The advantage to this approach is that the Laplace transformation converts the differential equation into an algebraic equation. This can usually be resolved easily according to the Laplace transform function $Y(s)$. Then, the Laplace inverse transform must be carried out in order to find the unknown function $y(x)$.

The Laplace transform $\mathcal{L}\{y(x)\}$ or image function $Y(s)$ of a function $y(x)$ can be found as follows:

$$Y(s) = \mathcal{L}\{y(x)\} = \int_0^\infty e^{-sx} y(x)\,dx,\ s \in \mathbb{R}.$$

In order to solve for the integral, the absolute value of the function must be exponentially bound. s merely acts as a variable in the transformation [3].

Properties

Laplace transforms of the derivatives $y'(x)$, $y''(x)$ are very crucial when solving differential equations.

$$\mathcal{L}\{y'(x)\} = s\,Y(s) - y(0)$$

$$\mathcal{L}\{y''(x)\} = s^2\,Y(s) - s\,y(0) - y'(0)$$

$$\mathcal{L}\{c_1 y_1(x) + c_2 y_2(x)\} = c_1\,\mathcal{L}\{y_1(x)\} + c_2\,\mathcal{L}\{y_2(x)\} \qquad \text{(linearity)}$$

$$\mathcal{L}\{y(ax)\} = \frac{1}{a}\,Y\left(\frac{s}{a}\right),\quad a > 0 \qquad \text{(Similarity theorem)}$$

$$\mathcal{L}\{e^{-ax}\,y(x)\} = Y(s + a),\quad a > 0 \qquad \text{(Complex shifting theorem)}$$

As you can see, the derivatives are converted into terms with the function $Y(s)$.

A few selected examples of the Laplace transforms of functions have been compiled in Table 14.

Table 14: Functions $y(x)$ and the corresponding Laplace transforms $Y(s)$

$y(x)$	$Y(s) = \mathcal{L}\{y(x)\}$
$y(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$	$\frac{1}{s}$
$y(x) = \begin{cases} 0 & x < 0 \\ x^{n-1} & x \geq 0 \end{cases}$	$\frac{(n-1)!}{s^n}\quad n = 1, 2, 3$
$y(x) = \begin{cases} 0 & x < 0 \\ \sin(ax) & x \geq 0 \end{cases}$	$\frac{a}{s^2 + a^2}$
$y(x) = \begin{cases} 0 & x < 0 \\ \cos(ax) & x \geq 0 \end{cases}$	$\frac{s}{s^2 + a^2}$
$y(x) = \begin{cases} 0 & x < 0 \\ e^{-ax} & x \geq 0 \end{cases}$	$\frac{1}{s + a}$

Fourier transform

During analysis, a time-dependent function is often found $f(t)$. Analysis is usually interested in the frequencies of the vibrations that underlie this function. For example, the undercarriage of a car should compensate for the bumps in the road so that the vehicle does not begin to rock back and forth. This requires a surface unevenness profile $h(x)$. If the car is driving down the road at a particular speed, it will begin to rock, which corresponds to the function $f(t)$. The characteristic frequencies must be determined from this function in order to be able to select and set the dimensions for the vehicle's shock absorbers, which will dampen the oscillations (see Undercarriage, Road excitation, Irregularity profiles).

The characteristic oscillation frequencies can be determined using the function $f(t)$, if this function $f(t)$ is subjected to a Fourier transform to form the function $F(\omega)$:

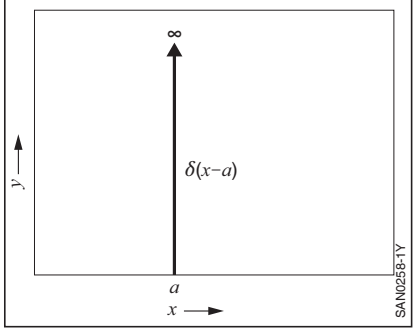
$$F(\omega) = \mathcal{F}\{f(t)\} = \int_{-\infty}^{\infty} f(t)\, e^{-i\omega t} dt,$$
where $i^2 = -1$.

This converts the time-dependent function $f(t)$ into the frequency-dependent spectral function $F(\omega)$ ([1], [2]).

This results in an inverse transformation along the lines of:

$$f(t) = \mathcal{F}^{-1}\{F(\omega)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega.$$

Figure 17: Symbolic representation of a Dirac delta function



In the literature, values of 1 (like here), $1/2\pi$ or $1/\sqrt{2\pi}$ are sometimes specified as the prefactor for the Fourier transform. This also changes the factor of the inverse transform. The critical aspect here is that the product of the transform's prefactors and the inverse transform yield $1/2\pi$.

Properties

Some properties of the Fourier transform are listed below.

$$\mathcal{F}\{c_1 f_1(t) + c_2 f_2(t)\} = c_1 \mathcal{F}\{f_1(t)\} + c_2 \mathcal{F}\{f_2(t)\}$$
(Linearity)

$$\mathcal{F}\{f_1(at)\} = \frac{1}{|a|} F\left(\frac{\omega}{a}\right), \quad a \in \mathbb{R},\ a \neq 0,$$
(Similarity theorem)

$$\mathcal{F}\{f(t - t_0)\} = e^{-i\omega t_0} F(\omega), \quad t_0 \in \mathbb{R},$$
(Shift theorem in the time range)

$$\mathcal{F}\{e^{i\omega_0 t} f(t)\} = F(\omega - \omega_0), \quad \omega_0 \in \mathbb{R},$$
(Shift theorem in the frequency domain)

$$\mathcal{F}\left\{\int_{-\infty}^{\infty} f(\tau)\, g(t-\tau)\right\} d\tau = \mathcal{F}\{f(t)\}\, \mathcal{F}\{g(t)\}.$$

Using the notation

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)\, g(t-\tau)\, d\tau = \int_{-\infty}^{\infty} f(t-\tau)\, g(t)\, d\tau$$

the convolution theorem can be summarized as follows:

$$\mathcal{F}\{(f * g)\} = F(\omega)\, G(\omega).$$

The transform in the frequency domain can be explained by the Dirac delta function, which is defined as follows [4]:

$\delta(t) = 0$ for $t \in \mathbb{R}$, $t \neq 0$, and

$$\int_{-\infty}^{\infty} \delta(t - t_0)\, f(t)\, dt = f(t_0).$$

This can also be written as

$$\delta(t - t_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(t-t_0)\tau}\, d\tau.$$

The Dirac delta function is always zero. However, at the point where its argument is zero, it can be said to be infinitely large. This is illustrated in Figure 17. The Fourier transform of the complex exponential function or the sine and cosine function with an angular frequency of ω_0 returns the Dirac delta function for ω_0:

$$\mathcal{F}\{e^{-i\omega_0 t}\} = 2\pi\ \delta(\omega - \omega_0)\ dx,$$

$$\mathcal{F}\{\delta(t - t_0)\} = \frac{1}{2\pi}\ e^{-i\omega t_0},$$

$$\mathcal{F}\{\sin(\omega_0 t)\} = i\pi\ [\delta(\omega + \omega_0) - \delta(\omega - \omega_0)],$$

$$\mathcal{F}\{\cos(\omega_0 t)\} = \pi\ [\delta(\omega + \omega_0) + \delta(\omega - \omega_0)].$$

For example, the Fourier transform is used to determine the frequencies of signals (e.g. the individual tones within the sound of a piano). In addition, the ratios of the corresponding amplitudes corresponding to the volume of sounds can be found. This allows sound waves to be broken down into individual sounds (frequencies) and their amplitude (volume) (see Undercarriage, Road excitation; Acoustics, Perceived sound levels).

Similarities and differenced between Laplace and Fourier transforms

The Laplace transform of a function $y(x)$

$$\mathcal{L}\{y(x)\} = \int_0^\infty e^{-sx} y(x)dx,\ s \in \mathbb{R}, \mathbb{C}$$

and the Fourier transform of $y(x)$

$$\mathcal{F}\{y(x)\} = \int_{-\infty}^{\infty} e^{-i\omega x}\ y(x)\,dx,\ \text{ where } i^2 = -1$$

are similarly defined. This means that both transforms have properties that are either similar or the same. Both transforms are linear, their similarity theorems are almost identical and the complex shifting theorem of the Laplace transform is equivalent to the shift theorem of the Fourier transform.

However, both transforms can be distinguished by their bounds of integration, because the Laplace integral begins at $x = 0$ and the Fourier integral at $-\infty$. In other words, the function $y(x)$ can be defined by for the entirety of $\mathbb{R}$ in the case of the Fourier integral, while $y(x)$ can only differ from zero for $\mathbb{R}^+$. Furthermore, the exponential function of the Laplace integral usually contains an arbitrary complex exponent $-s\,x$, while this term $-i\,\omega\,x$ is normally completely imaginary in the case of a Fourier integral.

The inverse of the Fourier transform can usually be calculated in a manner similar to that of the Fourier transform. However, calculating the inverse Laplace transform by means of complex curvilinear integrals is often difficult [1]. Tables such as Table 14 are usually helpful when solving differential equations in order to determine the original function $y(x)$ from the Laplace transform.

The direct correlation between the Laplace and the Fourier transform is depicted as follows: For the function $y(x)$, the function $g(x)$ is defined as:

$$g(x) = \begin{cases} 0 & x < 0 \\ e^{-\lambda x}\, y(x) & x \geq 0 \end{cases} \qquad \lambda \in \mathbb{C}$$

Subjecting $g(x)$ to a Fourier transform yields

$$\mathcal{F}\{g(x)\} = \int_{-\infty}^{\infty} e^{-i\omega x}\ g(x)\ dx$$
$$= \int_0^{\infty} e^{-i\omega x}\ e^{-\lambda x}\ y(x)\,dx.$$

Solving for $s = \lambda + i\omega$ then yields

$$\mathcal{F}\{g(x)\} = \int_{-\infty}^{\infty} e^{-i\omega x} g(x)dx = \int_{-\infty}^{\infty} e^{-sx} y(x)dx$$
$$= \mathcal{L}\{y(x)\}.$$

Matrix calculus

Vectors can indicate the intensity and direction of forces or of a velocity (see Vectors). In the case of a circular motion this can be easily represented using a vector, which specifies the initial direction of motion, together with a matrix, which describes the rotation.

Furthermore, matrices play the central role in solving systems of linear equations which occur in many technical questions. In order for example to calculate mechanical or thermal stresses in a component, complicated nonlinear differential equations are numerically solved. The standard method used for this purpose is the finite-element method (FEM), which ultimately (in very small time and space intervals) converts the differential equations into systems of linear equation and consequently into matrices. Reference is also made here by way of example to the use of matrices in the image processing of camera data in passenger cars (see Computer vision).

Firstly, matrices will be introduced in the following. Then arithmetical operations with matrices such as addition and multiplication will be defined. To conclude, three solution methods for systems of linear equation will be presented.

Structure of matrices
An $(m \times n)$ matrix A is made up of m rows and n columns. The component (elements, entries) a_{ik} in the ith row and kth column is a real or a complex number.

$$A = \begin{bmatrix} a_{11} & a_{12} \dots a_{1k} \dots a_{1n} \\ a_{21} & a_{22} \dots a_{2k} \dots a_{2n} \\ \vdots & \vdots \quad \vdots \quad \vdots \\ a_{i1} & a_{i2} \dots a_{ik} \dots a_{in} \\ \vdots & \vdots \quad \vdots \quad \vdots \\ a_{m1} & a_{m2} \dots a_{mk} \dots a_{mn} \end{bmatrix}$$

If the matrix has the same number of rows and columns ($m = n$), this is called a square matrix. The standouts among these are the upper and lower triangular matrices, the diagonal matrices, and the unit matrix. An upper or lower triangular matrix (A_o or A_u) has different entries on the diagonal and above or below zero; below or above the diagonals there are only zeros.

$$A_o = \begin{bmatrix} a_{11} & a_{12} \dots a_{1n} \\ 0 & a_{22} \dots a_{2n} \\ & \vdots \quad \vdots \\ 0 & 0 \dots a_{nn} \end{bmatrix}, \quad A_u = \begin{bmatrix} a_{11} & 0 \dots 0 \\ a_{21} & a_{22} \dots 0 \\ & \vdots \quad \vdots \\ a_{n1} & a_{n2} \dots a_{nn} \end{bmatrix}$$

A diagonal matrix D has different elements a_{kk} only on the diagonal. The unit matrix E is a diagonal matrix with ones as entries.

$$D = \begin{bmatrix} a_{11} & 0 \dots 0 \\ 0 & a_{22} \dots 0 \\ & \vdots \quad \vdots \\ 0 & 0 \dots a_{nn} \end{bmatrix}, \quad E = \begin{bmatrix} 1 & 0 \dots 0 \\ 0 & 1 \dots 0 \\ & \vdots \quad \vdots \\ 0 & 0 \dots 1 \end{bmatrix}$$

Arithmetical rules for matrices
Two matrices A and B are identical when they match in all components:
$A = B \Leftrightarrow a_{ik} = b_{ik}$ for all i, k.

Two matrices can only then be added or subtracted when they match both in their number of rows and their number of columns. In this case, the addition (subtraction) is performed for each component:
$C = A + B$, $c_{ik} = a_{ik} + b_{ik}$.

If a matrix is multiplied by a real or complex number λ, each individual components is multiplied by this scalar:
$B = \lambda A$, $b_{ik} = \lambda a_{ik}$.

This results in the following arithmetical rules for matrices:
$A + B = B + A$,
$A + (B + C) = (A + B) + C$,
$\lambda(\mu A) = (\lambda\mu) A$,
$(\lambda + \mu) A = \lambda A + \mu A$,
$\lambda(A + B) = \lambda A + \lambda B$.

A column vector x with n entries can be multiplied from the right by an $(m \times n)$-matrix as follows:

$$A\,x = \begin{bmatrix} a_{11} & a_{12} \ldots a_{1n} \\ a_{21} & a_{22} \ldots a_{2n} \\ & \vdots \quad \vdots \\ a_{m1} & a_{m2} \ldots a_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{bmatrix}$$

If the rows of matrix A are considered as vectors, i.e. the kth row corresponds to the row vector $a_k = (a_{k1}\ a_{k2}\ a_{k3} \ldots a_{kn})$, then in this multiplication in each case the scalar product is formed from the row vector a_k and the column vector x.

Thus, for the product of an $(m \times n)$ matrix with an n-row vector, an m-row vector is produced. This multiplication is linear, i.e. for matrix A, the two vectors x, y, and the two scalars λ, μ the following applies: $A(\lambda x + \mu y) = \lambda A x + \mu A y$.

Similarly, the multiplication of a row vector can be defined from the left with a matrix; this will not be discussed here.

Two matrices can be multiplied by each other when the number of columns s of the first matrix A matches the number of rows of the second matrix B. When the columns of matrix B are interpreted as vectors, the matrix multiplication is a natural extension of the multiplication of a matrix by a vector:

$$C = AB = \begin{bmatrix} a_{11} & a_{12} \ldots a_{1s} \\ a_{21} & a_{22} \ldots a_{2s} \\ & \vdots \quad \vdots \\ a_{i1} & a_{i2} \ldots a_{is} \\ & \vdots \quad \vdots \\ a_{m1} & a_{m2} \ldots a_{ms} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} \ldots b_{1j} \ldots b_{1n} \\ b_{21} & b_{22} \ldots b_{2j} \ldots b_{2n} \\ \vdots & \vdots \quad \vdots \quad \vdots \\ b_{s1} & b_{s2} \ldots b_{sj} \ldots b_{sn} \end{bmatrix}$$

$$= \begin{bmatrix} c_{11} \ldots c_{1j} \ldots c_{1n} \\ c_{i1} \ldots c_{ij} \ldots c_{in} \\ \vdots \quad \vdots \quad \vdots \\ c_{m1} \ldots c_{mj} \ldots c_{mn} \end{bmatrix}$$

where

$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} \cdots a_{is}b_{sj}$.

Similarly to the above, the new element c_{ij} corresponds to the scalar product from the ith row vector $a_i = (a_{i1}\ a_{i2}\ a_{i3} \ldots a_{is})$ of matrix A with the jth column vector $b_j = (b_{1j}, b_{2j}\ b_{3j} \ldots b_{sj})$ of matrix B. Thus, for the product of an $(m \times s)$ matrix with an $(s \times n)$ matrix an $(m \times n)$ matrix is produced.

The following arithmetical rules are obtained with regard to multiplication:

$A(BC) = (AB)C$,
$A(B+C) = AB + AC$,
$(A+B)C = AC + BC$,
$AE = EA = A$,

It must be noted in this connection that when multiplying two matrices as a rule the order may not be mixed up (not commutative).

Determinants

Square matrices can be characterized by the so-called determinant. When the columns (or rows) of the determinant are interpreted as vectors, the value of the determinant indicates the volume of the parallelepiped spanned by these vectors. It is clear from this that the determinant is only then different from zero when the column or row vectors are linearly independent.

Notations for the determinants

$$\det(A) = |A| = \begin{vmatrix} a_{11} & a_{12} \ldots a_{1n} \\ a_{21} & a_{22} \ldots a_{2n} \\ \vdots & \vdots \quad \vdots \\ a_{n1} & a_{n2} \ldots a_{nn} \end{vmatrix}$$

For (2×2) and (3×3) matrices the determinant can be calculated using Sarrus' rule. Here all the elements which are on diagonals going from top left to bottom right are multiplied by each other. These diagonals are then added. The elements which are on diagonals going from top right to bottom left are also multiplied by each other and subtracted from the other values. To this end, it is appropriate particularly for (3×3) matrices to write the first two column vectors on the right next to the determinant to simplify the calculation.

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}\,a_{22} - a_{12}\,a_{21}$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{matrix} a_{11}\,a_{22}\,a_{33} + a_{12}\,a_{23}\,a_{31} \\ + a_{13}\,a_{21}\,a_{32} - a_{13}\,a_{22}\,a_{31} \\ - a_{12}\,a_{21}\,a_{33} - a_{11}\,a_{23}\,a_{32} \end{matrix}$$

When the determinants have more than three rows (columns), the determinant can be calculated with the Laplace expansion.

For the unit matrix the following applies:
$\det(E) = 1$.

For the product of two square matrices the following applies:
$\det(A\,B) = \det(A) \cdot \det(B)$.

Inverse matrix
The square $(n \times n)$ matrix B is the inverse to the square $(n \times n)$ matrix A when the following applies:
$A\,B = B\,A = E$.

In this special case the multiplication is commutative. Then $B = A^{-1}$ is written and the preceding equation becomes
$A\,A^{-1} = A^{-1}\,A = E$.

The following is obtained for the determinant of an invertible matrix:
$\det(A) \neq 0$ and $\det(A^{-1}) = \dfrac{1}{\det(A)}$

Systems of linear equations
In the case of a system of linear equations real (or complex) numbers a_{ij} and b_j ($j = 1 \ldots m$) are given while the x_i ($i = 1 \ldots n$) are unknown and thus are to be determined.

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots \quad \cdots \quad \cdots \quad \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

These equations can be summarized compactly with the aid of the $(m \times n)$ matrix A and the two vectors b, x

$$A = \begin{bmatrix} a_{11} & a_{12} \ldots a_{1n} \\ a_{21} & a_{22} \ldots a_{2n} \\ \vdots & \vdots \\ a_{m1} & a_{m2} \ldots a_{mn} \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

as a matrix-vector equation:
$A\,x = b$

Three methods which can be used to solve the system of linear equations are set out below.

1) In the case of a square and furthermore invertible matrix A the following is obtained:
$x = A^{-1}\,b$.

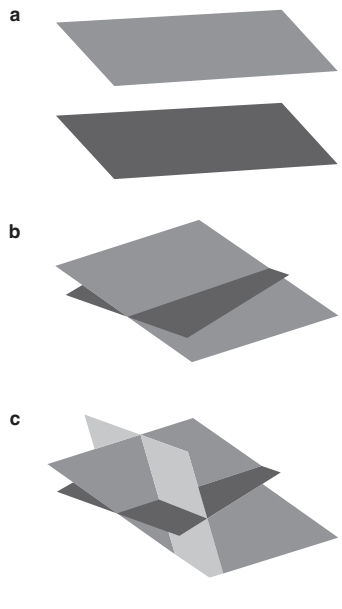
2) Alternatively, Cramer's rule an be used to solve a system of linear equations of a square and furthermore invertible matrix A. Here the component x_i of the solution vector x is obtained as:

$$x_i = \frac{\det(A_i)}{\det(A)}.$$

In the calculation of the auxiliary determinant A_i the $i$the column of the matrix is replaced by the vector b.

3) In the case of any $(m \times n)$ matrix the Gauss-Jordan algorithm can be used to determine the solution of the system of linear equations. Here matrix A is successively converted by row operations into an upper triangular matrix. The following are permitted as row operations:
– the multiplication of a row by a scalar (without zero), and
– the addition or subtraction of two rows.

Figure 18: Intersection of planes in space
a) Parallel planes,
b) Two planes intersect on a straight line,
c) Three planes intersect at a point.



Geometric interpretation of a system of linear equations

Each row of a system of linear equations can be interpreted geometrically as a hyperplane. Thus, the approach when solving a system of linear equations is to look for the points where these hyperplanes intersect. This is clearly shown in Figure 1 by way of example at two or three planes in $\mathbb{R}^3$: When the planes are parallel to each other, they do not intersect. Two planes intersect on a straight line. The latter corresponds exactly to a solution of the system of linear equations.

References

[1] T. Arens et al.: Mathematik. 3rd Ed., Verlag Springer Spektrum, 2015.
[2] I. N. Bronstein et al.: Taschenbuch der Mathematik. 10th Ed., Europa Lehrmittel, 2016.
[3] L. Papula: Mathematik für Ingenieure und Naturwissenschaftler, Volumes 1–3. Verlag Springer Vieweg, 2014.
[4] F. W. J. Oliver et al. (Editors): NIST Handbook of Mathematical Functions. 1st Ed., Cambridge University Press, 2010.

Technical statistics

Several thousand parts are required to produce an automobile. Since (random or systematic) fluctuations occur in every production process, not only the size for example of each part but also the associated tolerance is specified.

Even the measurement process is as a rule subject to random fluctuations, which are referred to as measurement uncertainty (previously the misleading term measuring error was also used). This measurement uncertainty should however be well below the tolerance limit of the part to be measured so that it can normally be disregarded.

Measurements play a central role in production and hence in this connection in quality assurance. Since the required failure rate must be very low for individual parts, i.e. often in the ppm range (parts per million), it is for the most part not possible for process-technical and economic reasons to measure all the parts. Instead, samples are taken in order then to infer the quality of the components with statistical methods. These statistical methods are presented in the following.

The aim of a production process is to work accurately and precisely. Accurate means that the mean value of the target figure is close to the specified value. Precise denotes the production when the dispersion (scatter) around the mean value is low. It is possible to check using the statistical methods set out in the following whether these requirements are satisfied.

Table 1: Linear measurement and associated absolute frequency of the measured value

l [mm]	190	191	192	193	194	195
h	0	0	27	23	0	0
l [mm]	196	197	198	199	200	201
h	0	0	34	0	0	2
l [mm]	202	203	204	205	206	
h	6	1	0	0	0	

Descriptive statistics

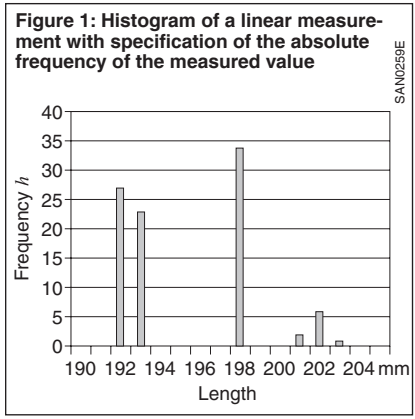
For the purpose of obtaining an overview of extensive data material, these data are summarized and illustrated in descriptive statistics in tables and graphics. In this way, already systematic errors (or even typing errors) can often be detected and eliminated. However, it is not possible with descriptive statistics to specify error probabilities, which can only be made on the basis of statistical parameter models of inductive statistics.

Absolute frequencies of a measurement

The starting point is a data set with n values $x_1, x_2, \ldots, x_n$ (e.g., linear measurements) on n different components. These measurements are summarized according to their characteristics, where for each length l_i that occurs the associated number (absolute frequency) h_i is compiled ($i = 1, 2, \ldots, k$; $k \leq n$). This is shown by way of example for a linear measurement in Table 1 and in the associated histogram (Figure 1).

In this example the length $l = 198$ mm occurs with the most frequency. For $l = 192$ mm and $l = 193$ mm an accumulation occurs, as well as around $l = 202$ mm. Linear values $l \geq 204$ mm and $l \leq 191$ mm were not measured in this sample.

Figure 1: Histogram of a linear measurement with specification of the absolute frequency of the measured value



The sample size n corresponds to the number of measurements and results as the total of the absolute frequencies:

$$n = h_1 + h_2 + \cdots + h_k = \sum_{i=1}^{k} h_i.$$

From this is obtained the relative frequency

$$f_i = \frac{h_i}{n} \text{ where } i = 1, 2, \ldots, k.$$

When the relative frequency is added up, this produces the value 1:

$$\sum_{i=1}^{k} f_i = 1.$$

To characterize a data set and to be able to compare this with other data sets, characteristic values such as the location parameters and dispersion parameters are calculated from this.

Location parameters and location measures

The median x_{med}, also called the central value, is the value in the middle of a data series $\{x_1, x_2, \ldots, x_n\}$ when the values are sorted by size. In other words, at least half of the data is less than or equal to x_{med} and at the same time at least half of the data is greater than or equal to x_{med}. In the case of an even number of data, x_{med} is the mean value of the two data in the middle of the series.

Often also of interest is the location of the lower (or upper) portion of values. The associated threshold value $x_{p\text{Q}} \in \{x_1, x_2, \ldots, x_n\}$ is also referred to as p quantile or as percentile. Hence, therefore, at least $p \cdot 100$ % of the data is less than or equal to $x_{p\text{Q}}$. At the same time, at least $(1 - p) \cdot 100$ % of the data is greater than or equal to $x_{p\text{Q}}$. The 25% quantile is also called the first or lower quartile; the 75 % quantile the upper or third quartile.

The arithmetic mean (average value) is defined as

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and the geometric mean of non-negative numbers x_i as

$$\bar{x}_{\text{geo}} = \sqrt[n]{x_1 \cdot x_2 \cdot \cdots \cdot x_n}.$$

The following always applies: $\bar{x}_{\text{geo}} \le \bar{x}$.

Likewise, the arithmetic and geometric means can also be calculated via the characteristics l_i and their frequency h_i.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} l_i h_i,$$

$$\bar{x}_{\text{geo}} = \sqrt[n]{l_1^{h_1} l_2^{h_2} \cdots l_k^{h_k}}.$$

The symbol μ is often used as well as the symbol $\bar{x}$ for the mean value.

Dispersion parameters

In a data set it is also important to know whether the individual values are close to or far from the mean value. This is recorded by the following dispersion parameters. The variance $\text{Var}(x) = s^2$ of the data set x is calculated as

$$s^2 = \text{Var}(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{k} h_i (l_i - \bar{x})^2.$$

The root from this is called standard deviation s.

$$s = \sqrt{s^2} = \sqrt{\text{Var}(x)}$$

σ is often used as well as s for the standard deviation.

Example

For the above example (Figure 1 and Table 1) the mean value is obtained at $\bar{x} = 195.40$ mm. The two frequencies $h_{192\text{ mm}}$ and $h_{193\text{ mm}}$ added up produce a number 50. This is above $n/2 = 46.5$. Thus, the median follows with $x_{\text{med}} = 193$ mm. The variance has the value $s^2 = 11.48$ mm^2 and the standard deviation $s = 3.38$ mm.

Probability calculus

The quality of a production process is to be inferred from a sample (e.g., linear measurement of a certain quantity of components). This is however only possible if realistic assumptions about the distribution of this measured value are made. A number of probability distributions are presented in the following for this purpose.

The Binomial distribution, the Hypergeometric distribution and the Poisson distribution are discrete distributions. Here, a sample is taken from a quantity and then the corresponding probabilities are calculated. From a mathematical standpoint both the sample size and the size of the quantity from which the sample is taken are natural numbers.

On the other hand, the Weibull distribution and the Gaussian normal distribution apply to continuous distributions, i.e. to real variables t or x.

Discrete distributions

Binomial distribution

A sample delivers only two values (Bernoulli experiment), one value with the probability p and the second value with the probability $q = 1 - p$ (since $p + q = 1 = 100\ \%$). An example in this connection is drawing a ball from a bowl containing balls of two different colors. The ball is returned to the bowl after each draw. Hence the probability for each removal of a sample is the same.

The associated probability function $f(x)$ indicates that this event in the case of n-fold execution occurs exactly x times:

$$f(x) = \binom{n}{x} p^x q^{n-x} = \binom{n}{x} p^x (1-p)^{n-x},$$

where $x = 0, 1, 2, \ldots, n$; $\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$.

The binomial distribution has the mean value $\mu = np$, the variance $\sigma^2 = npq = np(1-p)$, and the standard deviation $\sigma = \sqrt{npq} = \sqrt{np(1-p)}$.

Hypergeometric distribution

This distribution often plays a role in quality and final inspections. Similarly to the previous bowl model, there are at the start m white and $n - m$ red balls in a bowl. In contrast to the binomial distribution, this distribution corresponds to drawing a ball without putting it back, was which is customary for a sample in production. Hence, with each sample removal, the probability of drawing a white ball changes.

If now k samples are removed, this produces the probability function $f(x)$ at:

$$f(x) = \frac{\binom{m}{x}\binom{n-m}{k-x}}{\binom{n}{k}},\ x = 0, 1, 2, \ldots, k.$$

The hypergeometric distribution has the mean value

$\mu = k\frac{m}{n}$, the variance

$$\sigma^2 = \frac{km(n-m)(n-k)}{n^2(n-1)}$$

and the standard deviation

$$\sigma = \sqrt{\frac{km(n-m)(n-k)}{n^2(n-1)}}.$$

For $k < 0.05\,n$ the easier-to-calculate binomial distribution can be used as a very good approximation for the hypergeometric distribution.

Poisson distribution

If a binomial distribution exists with a very low probability of occurrence p, this can also be described by a Poisson distribution with the mean value μ and the variance $\sigma^2 = \mu$:

$$f(x, \mu) = \frac{\mu^x}{x!} e^{-\mu},\ x = 0, 1, 2, \ldots$$

The Poisson distribution has the mean value μ, the variance $\sigma^2 = \mu$, and the standard deviation $\sigma = \sqrt{\mu}$.

A binomial distribution can be well approximated by a Poisson distribution if the following two conditions are satisfied:
$np < 10$,
$n > 1500p$.

Continuous distributions
The Weibull distribution and the Gaussian normal distribution are classed as the continuous distributions. Hence they are functions of real variables x. The following functions are used to describe these: compression functions $f(x)$ and the associated distribution functions $F(x)$, which constitute their antiderivatives:

$$\frac{d}{dx}F(x) = f(x).$$

The integral over the compression function of x_1 through x_2

$$\int_{x_1}^{x_2} f(x)\,dx = F(x_2) - F(x_1)$$

specifies the probability that the random variable x is in the interval of x_1 through x_2. Accordingly, the distribution function

$$F(x_2) = \int_{-\infty}^{x_2} f(x)\,dx$$

is the probability that an event with $x \le x_2$ occurs.

Often the expectation value $E(x)$ is referred to instead of the mean value μ

$$\mu = E(x) = \int_{-\infty}^{\infty} x f(x)\,dx.$$

The variance is defined as

$$\sigma^2 = \mathrm{Var}(x) = E\left((x-\mu)^2\right) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx.$$

Weibull distribution
The fatigue of a material and the failure of a machine can often be described via a Weibull distribution over the time t (instead of the variable x) (service life T of the material or the machine, empirical or specific parameters α, β).

Accordingly, the probability that the service life ≤ t is at

$$F(t) = 1 - e^{-\alpha t^{\beta}}$$

and hence the survival probability (reliability)

$$R(t) = 1 - F(t) = e^{-\alpha t^{\beta}}.$$

The associated compression function $f(t)$ is the first derivation of $F(t)$:

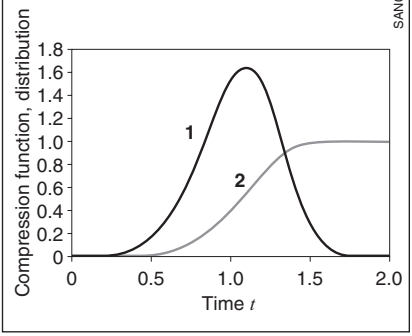
$$f(t) = F'(t) = \alpha\,\beta\,t^{\,\beta-1}\,e^{-\alpha t^{\beta}}$$
for $t > 0$, otherwise $f(t) = 0$.

From this the failure rate is obtained at

$$\lambda = \frac{f(t)}{R(t)} = \alpha\,\beta\,t^{\,\beta-1}.$$

Figure 2 shows a compression function and the associated distribution for $\alpha = 0.5$ and $\beta = 5$. The greatest failure probability is at $t = 1.1$. With increasing time the failure probability does in fact decreases again, but the function $F(t)$, i.e. the total failure rate until time t, naturally approaches 1 = 100 %.

Figure 2: Weibull distribution
For $\alpha = 0.5$ and $\beta = 5.0$.
1 Compression function,
2 Distribution.



Gaussian normal distribution
Very many random processes in nature and also in technology can be described by a Gaussian normal distribution. The associated compression function contains the expectation value μ and the standard deviation σ:

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

The associated Gaussian normal distribution function

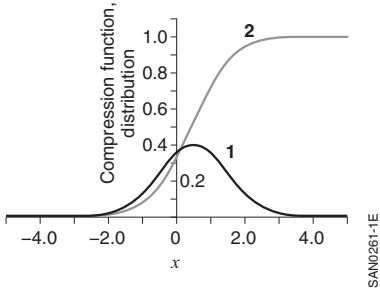
$$F(x,\mu,\sigma) = \int_{-\infty}^{x} f(z,\mu,\sigma)dz = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} e^{-\frac{(z-\mu)^2}{2\sigma^2}}dz, \quad -\infty < x < \infty$$

is the antiderivative for $f(x,\mu,\sigma)$. As the name already suggests, this is normalized, i.e.

$$\lim_{x\to\infty} F(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-\frac{(z-\mu)^2}{2\sigma^2}}\,dz = 1.$$

Figure 3 shows by way of example a Gaussian compression function and the associated Gaussian normal distribution. The compression function has its maximum at the expectation value; its breadth is determined by the standard deviation. Furthermore, it is axisymmetric to $x=\mu$. This "Gaussian bell-shaped curve" drops markedly for large x values and for small x values and approaches asymptotically the x axis.

Figure 3: Gaussian normal distribution
1 Compression function with the expectation value $\mu = 0.5$ and the standard deviation $\sigma^2 = 1$,
2 Associated normal distribution for the compression function.



The Gaussian normal distribution increases from 0 (for $x \to -\infty$) to the value 1 (for $x \to \infty$). At $x=\mu$ it has the value $F(x=\mu)=0.5$.

The Gaussian normal distribution has the mean value μ, the variance σ^2, and the standard deviation σ.

With the substitution $y = \frac{x-\mu}{\sigma}$ the following standard normal distribution is obtained

$$\varphi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}, \quad -\infty < z < \infty$$

Accordingly, the following normal distribution is obtained from this

$$\Phi(y) = \int_{-\infty}^{y} \varphi(z)\,dz = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{1}{2}z^2}\,dz, \quad -\infty < y < \infty.$$

Many production processes are characterized by way of their σ level (Table 2). In the case of the 3σ level, for instance, the probability that the produced parts are defect-free is 93.3 %.

The 6σ level (Six Sigma Level) plays a central role in quality management. Departing from the above definition of the compression function, long-term mean shifts are included here. Thus, the 6σ level corresponds to 3.4 DPMO (Defects Per Million Opportunities), i.e. $3.4 \cdot 10^{-6}$ of the parts are outside the tolerance band.

For a normally distributed random variable X with the mean value μ and the standard deviation σ the following probabilities are obtained:

Table 2: σ level
The middle column specifies the probability that the values are inside the range $\mu - k\sigma < x < \mu + k\sigma$.

σ level	inside [%]	outside [%]
σ	32	68
2σ	69	31
3σ	93.3	6.7
4σ	99.38	0.62
5σ	99.977	0.023
6σ	99.99966	0.00034

$$P(X \le x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \Phi(y).$$

$$P(X > x) = 1 - P(X \le x) = 1 - \Phi(y).$$

$$P(a \le X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

$$P(|X-\mu| \le k\sigma) = P(\mu - k\sigma \le X \le \mu + k\sigma) = 2\,\Phi(k) - 1.$$

Parameter estimation in the Gaussian normal distribution

Quantitative statements based on samples are to be made with the aid of statistical models (distribution models). The aim would be to determine characteristic values such as, for example, the mean value μ and the standard deviation σ. But this is not exactly possible since the sample is limited. However, it is possible to calculate from the sample estimates for the mean value and the standard deviation and for this to determine confidence intervals with which the sought-after characteristic value is in this interval with a specified probability.

To this end a random sample is taken from the sample size n, delivering the values $x_1, x_2, \ldots, x_n$. These elements are also called the realization of the sample. Calculated from these are the estimate of the mean value

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and the estimate of the variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Parameter estimations are carried out in the following, where it is assumed that there is an underlying Gaussian normal distribution. This distribution underlies very many production processes. For the parameter estimations for other distributions, reference is made to the bibliography ([3], [4], [5]).

Confidence interval for unknown mean value and known variance

In the case of a specified confidence level γ (usually 95 % or 99 %) the confidence interval of the breadth Δx in which the unknown mean value μ lies with a probability of γ is to be determined, i.e.

$P(|\bar{x} - \mu| \le \Delta x) = \gamma.$

This is equivalent to the unknown mean value μ lying in the interval

$\bar{x} - \Delta x \le \mu \le \bar{x} + \Delta x$

with a probability of $\gamma \cdot 100$ %. The estimate of the mean value is calculated from the sample.

With the standardized random variable

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \text{ and } \Delta z = \frac{\Delta x}{\frac{\sigma}{\sqrt{n}}}$$

the above probability can be depicted as follows:

$P(|z| \le \Delta z) = \gamma.$

From

$$P(|z| \le \Delta z) = 2\,\Phi(\Delta z) - 1 \Rightarrow \Phi(\Delta z) = 0.5\,(\gamma + 1)$$

the right side can be calculated and then Δz for example viewed in a table for the normally distributed Gaussian function, and finally from this Δx calculated.

Confidence interval for unknown mean value and unknown variance

From the sample first the estimates of the mean value $\bar{x}$ and the standard deviation s are calculated. Similarly to before, in the case of a specified confidence level γ (usually 95 % or 99 %) the confidence interval of the breadth Δx in which the unknown mean value μ lies with a probability of γ is to be determined, i.e.

$$P(|\bar{x} - \mu| \le \Delta x) = \gamma.$$

The procedure for determining Δx is performed exactly as before, where for σ now s is used here.

Sums and products of random variables

If a process is determined by a number of random variables $x_1, x_2, \ldots, x_n$, this produces for the sum total $z_S = x_1 + x_2 + \cdots + x_n$ and for the product $z_P = x_1 \cdot x_2 \cdot \cdots \cdot x_n$ the expectation values E and the variance:

$E(z_S) = E(x_1) + E(x_2) + \cdots + E(x_n)$
or $\mu_{zS} = \mu_1 + \mu_2 + \cdots + \mu_n$.

$E(z_P) = E(x_1) \cdot E(x_2) \cdot \cdots \cdot E(x_n)$
or $\mu_{zP} = \mu_1 \cdot \mu_2 \cdot \cdots \cdot \mu_n$.

$\mathrm{Var}(z_S) = \mathrm{Var}(x_1) + \mathrm{Var}(x_2) + \cdots + \mathrm{Var}(x_n)$
or $\sigma_{zS}^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$.

Linear regression

In the case of a sample a linear connection between the parameter x and the measured variable y is presumed:

$$y = mx + t.$$

The sample delivers for the set parameters x_i the measured values y_i (i = 1, 2, …, n). From these values the gradient of the straight line m and the axis intercept t are to be determined so that the deviation between the straight line and the measured values (x_i, y_i) is as small as possible (minimum Gaussian deviation). These are obtained then as

$$t = \frac{\left(\sum_{i=1}^{n} x_i\right) \cdot \left(\sum_{i=1}^{n} x_i y_i\right) - \left(\sum_{i=1}^{n} x_i^2\right) \cdot \left(\sum_{i=1}^{n} y_i\right)}{\left(\sum_{i=1}^{n} x_i\right)^2 - n \sum_{i=1}^{n} x_i^2}$$

and as

$$m = \frac{\left(\sum_{i=1}^{n} x_i\right) \cdot \left(\sum_{i=1}^{n} y_i\right) - n \sum_{i=1}^{n} x_i y_i}{\left(\sum_{i=1}^{n} x_i\right)^2 - n \sum_{i=1}^{n} x_i^2}$$

In the case of a linear connection $y = mx$ without axis intercept, m is simplified at

$$m = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}.$$

Characteristic value in quality assurance and in quality management

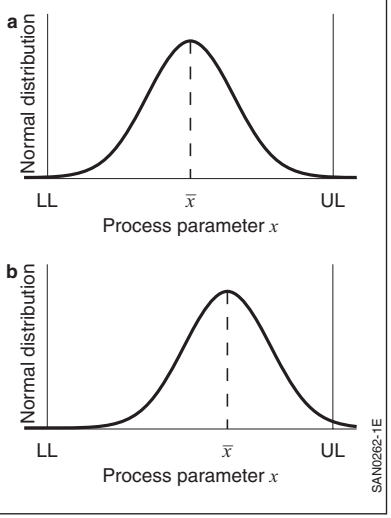
Statistical Process Control (SPC) is used in production processes and in quality management to assure the quality of the manufactured goods. A Gaussian normal distribution is assumed as a rule here.

In a production process the upper and lower limits (UL and LL) are stipulated for example by customer requirements. The tolerance range T is obtained from this as

$$T = \mathrm{UL} - \mathrm{LL}.$$

The mean value $\bar{x}$ and the standard deviation s are calculated from a sample. Furthermore, the critical length Z_{crit} is determined, i.e. the amount of the difference between the mean value $\bar{x}$ and UL or between the mean value $\bar{x}$ and LL is formed, and the smaller of these is taken as Z_{crit}.

Figure 4: Gaussian distributions of two samples
a) Mean value lies exactly in the middle of the lower and upper limits,
b) Mean value is closer to the upper limit.
$\bar{x}$ Mean value,
LL Lower limit,
UL Upper limit.



$Z_{crit} = \min(|UL - \bar{x}|, |LL - \bar{x}|)$.

It can be seen in Figure 4 that the mean value is closer to the UL than to the LL. Hence here
$Z_{crit} = UL - \bar{x}$.

From this the process capability c_p is obtained as

$$c_p = \frac{T}{6s}$$

and the process capability characteristic value c_{pk} as

$$c_{pk} = \frac{Z_{crit}}{3s}.$$

In standard production $c_p \geq 1.33$ and $c_{pk} \geq 1.33$ are classed as guide values for a stable process which satisfies the quality requirements.

References
[1] T. Arens et al.: Mathematik. 3rd Ed., Verlag Springer Spektrum, 2015.
[2] I.N. Bronstein et al.: Taschenbuch der Mathematik. 10th Ed., Europa Lehrmittel Verlag, 2016.
[3] L. Papula: Mathematik für Ingenieure und Naturwissenschaftler, Volumes 1 to 3. 14th Ed., Verlag Springer Vieweg, 2014
[4] A. Steland: Basiswissen Statistik. 3rd Ed., Verlag Springer Spektrum, 2013.
[5] G. Bamberg, F. Baur, M. Krapp: Statistik. 17th Ed., Oldenbourg Verlag, 2012.

Finite-element method

Applications

Virtually all technical procedures can be simulated on a computer with the finite-element method FEM (the term having been introduced by Ray W. Clough at the start of the 1960s [1]). However, this involves breaking down any body (gaseous, liquid, or solid) into elements that are simple in shape (line, triangle, square, tetrahedron, pentahedron, or hexahedron), that are as small as possible, and that are permanently connected to each other at their corner points ("nodes"). Small elements are important because the behavior of elements formulated by approximation using linear equations is only applicable to infinitesimal elements. However, the computing time calls for finite elements. The approximation to reality is better the smaller the elements are.

The application of FEM in practice – also known as FEA (finite-element analysis) – began in the early 1960s in the aviation and aerospace industries and followed soon after in automobile manufacturing. Today the method is used in all fields of technology, including weather forecasting, medical science, and for many sectors of automobile manufacturing ranging from engine and chassis components through to body calculations and crash behavior.

There are two different types of application. Firstly, virtually fully automatic "Black Box" FEM contained in all CAD programs (computer-aided design) for rough calculations by the design engineer (e.g. in designing a bumper) and, secondly, the use reserved for specialists of special FEM programs (e.g. in body calculations, in axle development or in driving dynamics).

FEM program system
The software of an FEM system consists of a preprocessor, a postprocessor, and the actual FEM program. Network creation, i.e. breakdown into elements, is mainly performed in the preprocessor on the basis of a CAD geometry which is read directly or via neutral interfaces such as IGES (Initial Graphics Exchange Specification), VDA-FS (Verband der Automobilindustrie – Flächenschnittstelle) or STEP (Standard for the Exchange of Product Model Data). The FEM program calculates the computing model formulated in this way. The result found is then shown in graphic form in the postprocessor (e.g. stress distribution by means of isocolors, deformations as motion animation).

Basic knowledge for application
FEM is, like all numerical methods, an approximation process. In mechanics, the main area of application, the limitations caused by this are described in the following.

Small motions in one solution step
Bodies move on paths which are normally higher-order curves. With the basic principle of linearization of all processes, this motion is limited to a straight path which can then be described by linear equations. When transferred to the element corners (nodes), they also move on a straight line. Thus, the nodes are only able to realize very small motions correctly (node twists less than 3.5°). The actual motion along any path or nonlinear material behavior is thus solved linearly with many small steps.

Calculation accuracy
The linear equation system is formulated and solved with the limited computing accuracy of a computer. Usually, 8 bytes (= 64 bits) are used with a computing accuracy of 13 significant digits for the number stored, i.e. only the first 13 digits of a number can be represented exactly. The further digits in this number are random numbers. As a result, this rules out the possibility of any stiffness differences of the individual components in a model. Therefore, in the deformation calculation of a body, it is necessary, as for example in the measurement of body deformation, to replace the axle springs with rigid supports.

Interpretation of results
The great danger lies in the fact that a formal computation model formulated correctly by a beginner will indeed deliver beautifully colorful images, but the results shown can center around factors far from reality.

The problems resulting from the above-mentioned limitations must therefore be identified and signaled by the computing program, so that the less experienced user is also able to obtain correct results easily.

Areas of FEM application
In technology terms, physics is generally divided into five areas – mechanics with statics and kinematics (e.g. body, axle), dynamics with acoustics (e.g. vehicle noise), thermodynamics (e.g. temperature distribution in the engine), electricity with magnetism (e.g. ignition coil, sensor technology), and optics (e.g. headlamp). When it comes to FEM, a distinction is always made between

– linear and nonlinear static and dynamic problems with the deformations as an unknown for stress calculation and dynamic analysis,
– stationary (timeless) and nonstationary (time-dependent) potential problems (e.g. temperature, sound pressure, electrical or magnetic potential) with the potentials as an unknown,
– and the linking of these different fields, e.g. to calculate a temperature field and the resulting deformations, stresses and forces in linear statics when the engine is started.

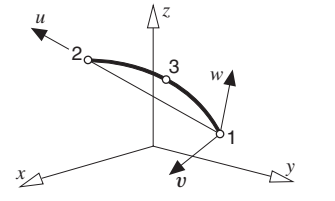
Elements of FEM
The properties of the elements define the most important performance data of an FEM program. The element quality is determined by the degree of the mathematical formulation function selected. Here, for example, a distinction is made between elements with a linear or quadratic formulation, recognizable from the midside node in the middle of the edge. The quality of a computing model is therefore dependent not just on the fineness of the used mesh, but also quite considerably on the formulation function.

A distinction is made between three different types of elements: line elements, shell elements, and volume elements.

Line elements
Line elements (Figure 1) are either straight or curved with an midside node. The cross-sections are described by specifying the numerical values for the cross-sectional area A, the reduced shear cross-sections $A_{\text{red-v-w}}$ (shear areas), the principal moments of inertia (I_v, I_w), the torsional moment of inertia (I_t) with the torsional section modulus (W_t), the sector moment of inertia for warping-force torsion, an angle α which describes the position of principal axes of inertia v and w in relation to the road plane, and the maximum four stress points (S_v, S_w) for bending-stress calculation.

Shell elements
Shell elements (Figure 2) are either triangular or quadrangular in shape – ideally an equilateral triangle or a square, usually of constant thickness. If the midside nodes are omitted, the edges are straight.

Figure 1: Line elements
With constant or linearly variable cross-section.



Volume (solid) elements
Volume elements (Figure 3) as tetrahedrons, pentahedrons and hexahedrons without midside nodes have straight edges – ideally an equilateral tetrahedron or a cube. With a sufficient number of elements in relation to the thickness (greater than or equal to three), elements with midside nodes are not necessary today. However, this does not apply to tetrahedrons, which are almost alway used in the case of a complicated geometry, such as a cylinder head for example.

Modeling and evaluation of results
The most important function in using an FEM program is the usually time-consuming task of creating the input data as a computing model with the preprocessor. The user should try to achieve this target with as few elements and nodes as possible (however, the computing model of a body today has approximately three to four million nodes). To do so, the user requires a certain level of experience, but must have an exact knowledge of the qualities of the elements used (see FEM examples). These can differ slightly in every FEM program.

The first step in modeling involves choosing the element type (line, shell or volume element) and determining the fineness of the mesh, e.g. by means of the specified middle element-edge length. The next step involves defining the properties (material data), e.g. the element thickness for unit elements, the cross-section values for line elements, and the units used (e.g. length in mm and force in N). A further step involves determining support conditions and load. The crucial factor here is considering the points where a model is fixed, and where it is loaded. In relation to loading, it is also useful to conduct a breakdown of the total load into load cases, e.g. into the mass load from gravity weight and different traffic loads.

All FEM results are available in list form or postprocessor data format, and can therefore be shown in graphic form (see FEM examples). To this end, the postprocessor offers all the conceivable forms of display.

Figure 2: Shell elements
a) Triangle, b) Quadrangular.

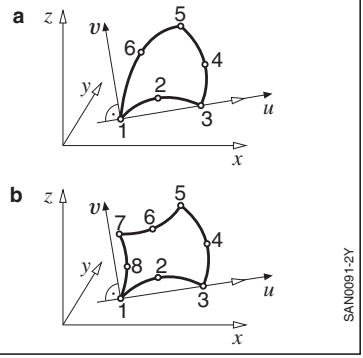
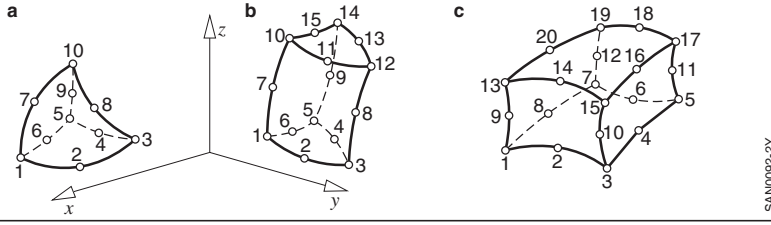


Figure 3: Volume elements
a) Tetrahedron, b) Pentahedron, c) Hexahedron.



FEM examples

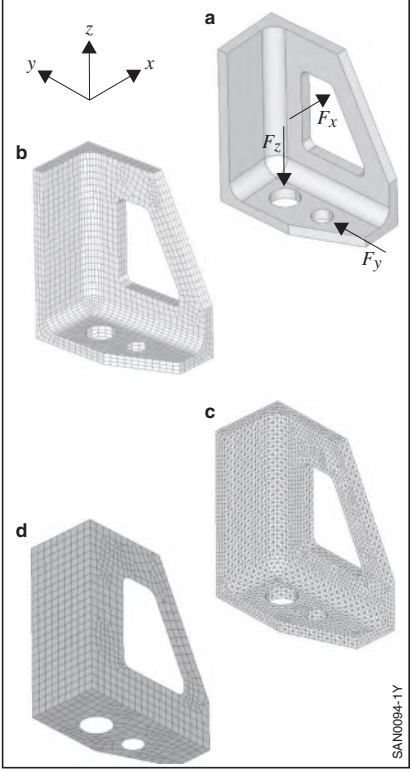
For the examples, modeling is performed on the basis of CAD geometry using the FEM program TP 2000. The model inputs, together with color representation of the results, can also be found on the Internet page specified in [2].

In reality, all bodies are three-dimensional. In simulation, a simplified solution is often chosen in order to save time and expense. It is much easier, for example, to realize the automatic meshing of a flat surface in shell elements than it is to realize a body in its volume elements. The frequently used tetrahedron mesh, which today every preprocessor creates for any volume geometry, does not always live up to expectations.

In automobile manufacturing, the structural components are either thick-walled and solid (e.g. engine, transmission, axles, wheels) and modeled with volume elements or they are made from thin-walled sheet metal (e.g. automobile body, truck cab) and are modeled with shell and line elements.

The first example, as a volume-element model, belongs to the first group of thick-walled, solid components. It is compared with the commonly used shell model of the second group. This also includes the line element commonly used in automobile manufacturing with the second example.

Figure 4: Cast-steel engine mount
a) CAD model with load F_x, F_y, F_z,
b) Model A1 (hexahedron),
c) Model B1 (tetrahedron),
d) Model C (shell).



Example 1: Cast-steel engine mount as shell- and volume-element model (linear statics)

The aim is to compare shell and volume elements in linear statics on a relatively thick-walled (3.75 mm) cast-steel engine mount ($50 \times 25 \times 57$ mm^3). In the volume elements, two models A and B (each with and without midside nodes) with significantly different result qualities are compared; added to this is shell model C (Figure 4). Starting out from the CAD volume geometry, the preprocessor automatically creates volume-element networks A and B, and also shell-element model C by means of the surface model.

The material properties in the units mm, N and kg are:
Modulus of elasticity: 210,000 N/mm^2
Poisson's ratio: 0.3
Density: 0.00000785 kg/mm^3 .

In the case of models A and B, the element properties are defined with the element type "Solid" (volume) with three node degrees of freedom v_x, v_y and v_z. In the case of model C, the properties are defined with the element type "Plate" (shell) with six node degrees of freedom v_x, v_y, v_z, d_x, d_y, d_z and constant thickness d = 3.75 mm.

In model A, the meshing with volume elements (solid mesh) shows the preferred breakdown into hexahedrons (not possible fully automated for all geometries because here, for example, the geometry

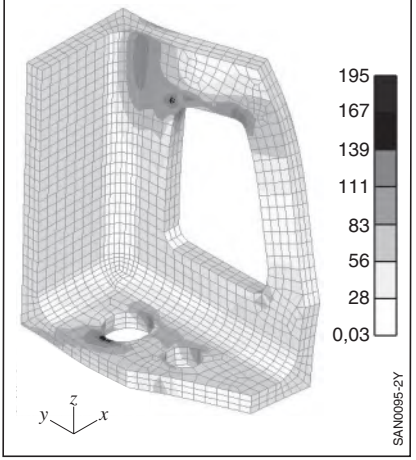
first had to be broken down into basic bodies) and in model B, the automatic tetrahedron meshing. The number of elements in relation to thickness which is crucial for accuracy is specified (at least three elements here).

Support conditions
On the rectangular xz plane, $v_y = 0$ applies to all nodes. All the edge nodes on the right long leg are pinned by $v_x = 0$, and those of the lower short leg are pinned by $v_z = 0$.

Loading
$\sum F_x = 900$ N, $\sum F_y = 2006$ N, $\sum F_z = -550$ N. All the loads are defined as surface load ("on surface", for shell "on curve") at the recess at $F_x = 600$ N, at the small hole at $F_y = 2006$ N, and at the large hole at $F_z = -550$ N.

Note: "Isolated" individual loads are only permitted with line elements.

Figure 5: Deformation and stress distribution in model A2 as the correct model
$v = 0.064$ mm, $\sigma = 195$ MPa.



Result
The results are shown in Table 1. Figure 5 shows the result for model A with stresses as shades of gray (critical areas are shown in dark) or as isocolors in the original.

Volume elements are very sensitive to incorrectly approximated load distributions. Volume elements should therefore only ever be subjected to surface loads.

Considering Table 1 in detail: Experience shows that model A2 with midside nodes delivers the correct result ($v = 0.064$ mm, $\sigma = 195$ MPa). The deviation between averaged and maximum node stress (from all the elements at the same node) should be as small as possible (less than 15 %). This is achieved with model A1 at 11 % with three elements per thickness.

In model B1 with three tetrahedrons per thickness, the deformation is 25 % too small ($v = 0.048$ mm), and the maximum stress is also 25 %. With only two tetrahedrons in relation to the thickness, the model would even be 58 % too rigid and can barely be used with correspondingly lower stresses. However, with midside nodes (B2), it delivers virtually identical deformations and stresses to A2 with a very large number of nodes and long computing times. Caution is therefore advised with rough tetrahedron mesh without midside nodes.

Table 1: Results for engine-mount model
Values in parentheses apply to elements with midside nodes.

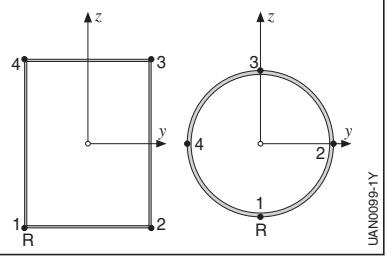
Model	Element type	No. of nodes	Weight in kg	Max. deformation in mm	Mean stress in MPa	Max. stress in MPa	Stress error max. in %	Computing time in sec
A1 (A2)	Solid hexahedron	4,000 (15,000)	0.119	0.061 (0.064)	145 (185)	164 (195)	15 (0)	10 (110)
B1 (B2)	Solid tetrahedron	10,000 (70,000)	0.119	0.048 (0.064)	115 (173)	146 (209)	30 (6)	60 (2,400)
C	Rectangular shell	766	0.114	0.081	197	223	14	2

Model C with shell elements without shear deformation (only intended for thin-walled structures) is clearly too soft (v = 0.081 mm, 21 % greater, as thick-walled with shear deformation even 30 % greater). Shell elements, whether thin- or thick-walled, deliver only partially usable deformations with relatively large thickness, particularly in the case of very compact bodies, as is the case here. However, at stresses of +14 %, it is for the most part on the safe side.

Example 2: Tubular body frame
The aim is to carry out an FEM analysis on a body in the form of a spaceframe without sheet paneling, including optimized weight and stiffness for the example of a mini-pickup (not a real vehicle). The material properties are:
Modulus of elasticity: 200,000 N/mm^2
Poisson's ratio: 0.3
Density: 0.00000785 kg/mm^3 .

For reasons of simplicity, only two shapes are used as profile sections (Figure 6). A box-type profile section (90 × 120 × 1.5 mm^3) and a tube profile section (70 x 2 mm^2), which are defined by their shape in the preprocessor in the properties directly with the element type "Bar" and its constant cross-section (different starting and end cross-sections would be "Beam"). This results in the required cross-section values as follows (in mm^2 or mm^4 and mm^3).

Figure 6: Profile sections of spaceframe
a) Box-type section, b) Tube section.



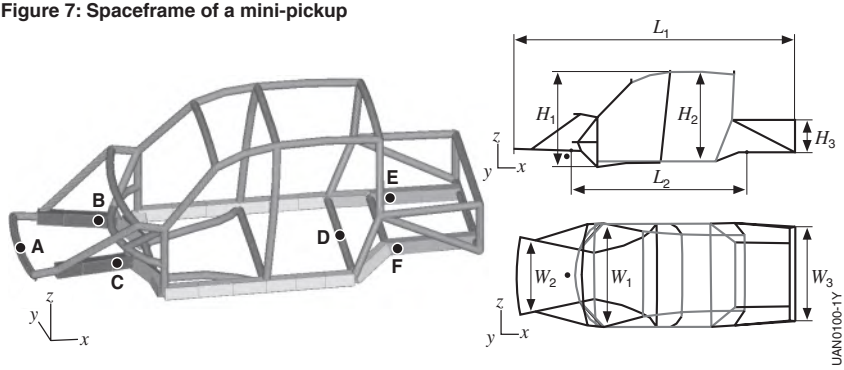
Box/tube (Figure 6)
– Cross-sectional area:
  A = 621/427.
– Reduced shear cross-sections (shear area):
  A_{redI} = 325/227,
  A_{redII} = 219/227.
– Principal moments of inertia:
  I_I = 1,348,306/247,168
  I_{II} = 869,596/247,168.
– Torsional moment of inertia:
  I_t = 1,606,083/494,261.
– Torsional section modulus:
  W_t = 7,334/2,177.
– Position of principal axes of inertia in relation to road:
  α = 0°.
– Maximum four stress points:
  S_x = –45/45/45/–45 / 0/35/0/–35,
  S_y = –60/–60/60/60 / –35/0/35/0.

The main dimensions according to the side and top views (Figure 7) are (in mm):
L_1 = 4,114 (max.)
L_2 = 2,650
W_1 = 1,517 (max.)
W_2 = 1,147 (front)
W_3 = 1,374 (rear)
H_1 = 1,402 (max.)
H_2 = 1,315
H_3 = 469 (box)

Body
The tubular consists of 18 components (Figure 8), e.g.:
2 Side pillar,
3 Pedal cross-pillar,
7 Cockpit cross-pillar,
8 Side pillar, front,
10 Auxiliary cross-pillar, front,
14 Roof frame, side,
15 Roof frame, lateral,
16 Rear subframe,
17 Side pillar, box,
19 Bumper, rear.

Support conditions
Linear statics, load case – bending (not shown):
A v_y = 0; (A – F, see Figure 7).
B, C, E, F v_z = 0; D v_x = 0, v_y = 0.

Figure 7: Spaceframe of a mini-pickup



Linear statics, load case – torsion:
A, B, C = free;
E, F $v_z = 0$;
D v_x, $v_y = 0$, $d_y = 0$, $d_z = 0$.

Free/free dynamics:
No supports, body vibrates freely in the springs.

Loading
Linear statics, load case – torsion:
Torsional moment 3,000,000 Nmm as regular unit load on
Front-axle bearings B: $F = -3593.7$ N
Front-axle bearings C: $F = 3593.7$ N

Free/free dynamics:
Only masses from gravity weight with 170 kg, no supplementary masses.

Results
Free/free dynamics:
– The body can vibrate freely in its springs, the lower eigenvalues and modes of the undamped, elastic system in ascending order; for free/free beginning with several rigid-body vibration modes (here 1 through 6) with the natural frequency 0 in each case.
– The 1st eigenvalue (38 Hz) is obtained as the 1st torsional vibration, the 3rd eigenvalue (48 Hz) as the 1st flexural vibration; with in each case the normalized deformations x, y, z at all nodes (here normalized to max. 0.1 mm), the normalized stresses in all elements, and the deformation processes important for weight optimization per element and per component 1 through 18 in % as a bar chart (as shown for linear statics in Figure 8). Absolute values are only obtained in the event of an excitation calculation.

Linear statics:
– Deformations x, y, z at all nodes,
– Reaction forces and moments at the support nodes, stresses at all elements, strain energy per element and per component in %.

Weight and stiffness optimization
The computation formula for weight and stiffness optimization has been known since the 1960s (error max. 10 % for doubled rigidity). The stiffness change of the overall structure (in %) is obtained from the stiffness change of the component multiplied by the deformation-process ratio of the component, divided by 100.

This formula is used specifically for optimization purposes in virtually all automobile-critical "Torsion" load cases with consideration of the other load cases. In this way, the structure can be reinforced using pillars that take most of the load, and reduce weight via the pillars that take less of the load.

Application of formula
1st component (see Figure 8):
Component 15 (roof frame, lateral, G =11.85 kg with 14.57 % strain energy ratio:
Produces (14.57 x 116) / 100 = 16.9 % change (i.e. reduction) of torsion between the axles when the stiffness (planar moment of inertia) of this component is increased by a factor of 2.169 (116.9 %). The weight increase here is 3.55 kg when the tube diameter is increased from 70 to 90 mm.

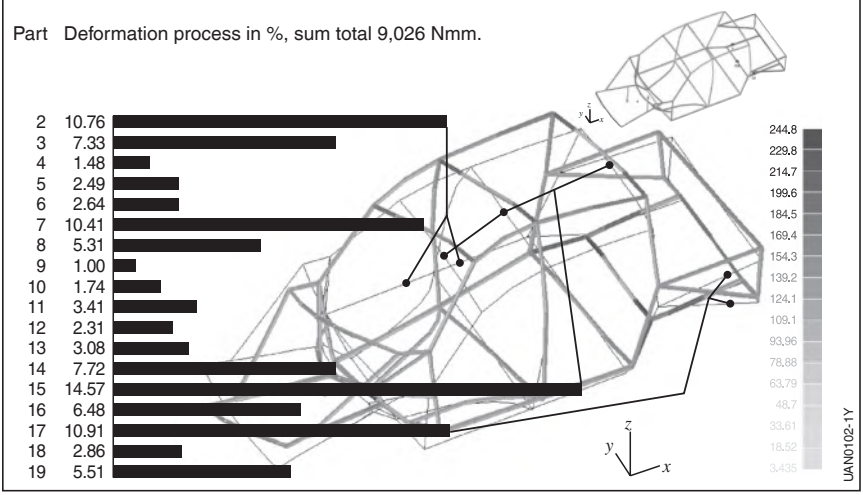
2nd component (see Figure 8):
Component 6 (linear reinforcement, G = 11.14 kg) with 2.64 % strain energy ratio:
Produces (2.64 x 250)/100 = 6.6 % change (i.e. only very small reduction of torsion between the axles when the stiffness (moment of inertia) of this component is reduced by a factor of 3.5 (250 %). The weight reduction is 3.34 kg when the tube diameter is reduced from 70 to 50 mm.

Result:
A minimal weight increase of 3.55 – 3.34 = 0.21 kg increases torsional stiffness by 16.9 – 6.6 = 10.3 % (checking with altered profile sections produces 9 %, this covers the low error in the computation formula).

A look at the strain energy charts of the other load cases (not pictured here) shows that this also applies to torsional vibration and is of no importance to bending. It is thus possible to significantly increase the torsional stiffness and flexural bending stiffness of this body still further and to safely reduce the total weight by reducing the cross-sections of oversized components.

References
[1] R.W. Clough: The Finite Element Method In Plane Stress Analysis; Publication, 2nd ASCE Conference on Electronic Computation,1960.
[2] www.IGFgrothTP2000.de.
[3] D. Braess: Finite Elemente: Theorie, Schnelle Löser und Anwendungen in der Elastizitätstheorie. 5th Edition, Springer Spectrum, 2013.

Figure 8: Load case, torsion: 3,000,000 Nmm at the front-axle points
Maximum deformation in mm:
x direction: 0.868459; y direction: 3.071005; z direction: 3.688961.
Maximum stress: 64 N/mm².



Control engineering

Terms and definitions

(in accordance with DIN 19226 [1])

Closed-loop control
In a technical process, the function of closed-loop control is to place and maintain a particular physical parameter (controlled variable y) at a specified value (e.g. the alternator voltage in the electrical system). Here, the controlled variable is continuously measured and compared with the specified value (the reference variable w, e.g. the setpoint input of the alternator voltage as a function of the state of charge of the battery) (Figure 1). In the event of a deviation, a suitable adjustment of the correcting variable u (e.g. influencing the excitation current in the alternator) on the process (controlled system) is carried out in such a way that the controlled variable y is again adapted to the specification. This process takes place in a closed control loop. Deviations can occur when disturbances z (e.g. the activation of a further electrical consumer) act on the controlled system and affect the controlled variable y in an undesirable way ([2], [3] and [4]).

Closed-loop control operations are performed at many points in a motor vehicle. Further examples include control of cooling-water temperature, A/C control, and many other control operations from the engine (knock control, λ control), transmission (clutch control) and chassis (yaw-rate control).

Figure 1: Basic structure of a control loop
y Controlled variable,
w Reference variable,
u Correcting variable,
z Disturbance.

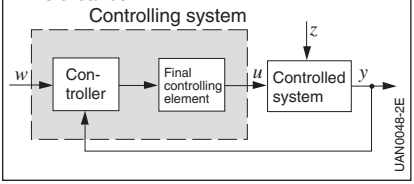
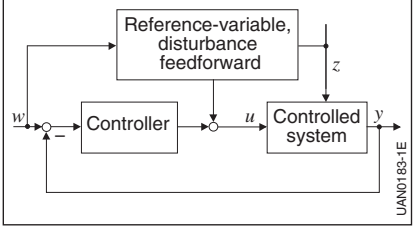


Figure 2: Control loop with open-loop-controlled functions (reference-variable and disturbance feedforward)
y Controlled variable,
w Reference variable,
u Correcting variable,
z Disturbance.



Open-loop control
Open-loop control is essentially also used instead of closed-loop control. In this case, the closed-loop controlling system is replaced by an open-loop controlling system and feedback of the controlled variable is omitted. This process is only then possible if the behavior of the controlled system is known exactly and if no (non-measurable) disturbances z are acting on the controlled system.

Open-loop control is preferred here as no stability problems can arise on account of there being no feedback. Because the previously mentioned preconditions are rarely satisfied in practice, the use of closed-loop control is for the most part unavoidable.

Combination of open- and closed-loop control
Open- and closed-loop control are, in practice, frequently combined in order to exploit the specific advantages of the two structures. Here, established interconnections between reference variable, disturbance, correcting variable and controlled variable are linked as much as possible in order to realize these as open-loop control. Deviations that still arise on account of changed parameters or non-measurable disturbances are corrected by closed-loop control (Figure 2).

Cascade control
A structure is frequently encountered in which the controlled system is separated into two or more partial systems (e.g. into the actual process and the associated actuator). On the basis of this separation, there are one or more internal controllers and one external controller, which are designed and adjusted separately. This procedure is referred to as cascade control (Figure 3).

The controller design is simplified by this separation of the control task into several manageable subtasks. There are additional advantages in the dynamic response due to the fact that disturbances which act in the internal control loop are corrected there before they affect the external control loop. This makes the entire control operation quicker. Likewise, nonlinear characteristic curves of the internal loop can be linearized.

Cascade control is used in many control systems in motor vehicles, e.g. in current control for electrohydraulic actuators or in position control for electric-motor actuators.

Control-engineering transfer elements

A control loop must satisfy four essential requirements in terms of performance:
– The control loop must be stable,
– The control loop must demonstrate a specific stationary accuracy,
– The response to a step change of the reference variable must be sufficiently damped,
– The control loop must be sufficiently fast.

To satisfy these partly contradictory requirements, it is first necessary to describe the static and dynamic responses

Figure 3: Cascade control loop
y_1, y_2 Controlled variables,
w_1, w_2 Reference variables,
u Correcting variable.

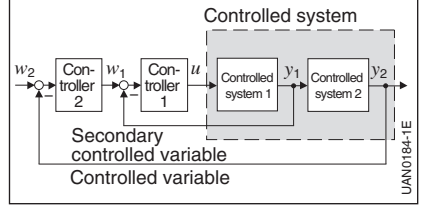
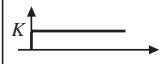
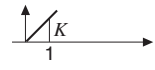
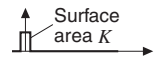
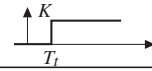
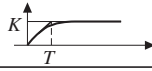
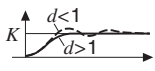


Table 1: Summary of important transfer elements

Transfer element	Differential equation (input u, output y)	Transfer function	Step response
P element	$y = Ku$	K	
I element	$y = K\int_0^t u(\tau)\,d\tau$	$\frac{K}{s}$	
D element	$y = K\dot{u}$	Ks	
Lag element	$y(t) = Ku(t-T_t)$	$Ke^{-T_t s}$	
P-T_1 element	$T\dot{y}+y = Ku$	$\frac{K}{1+Ts}$	
P-T_2 element	$T^2\ddot{y}+2dT\dot{y}+y = Ku$	$\frac{K}{1+2dTs+T^2s^2}$	

of the control-loop elements (controlled system and controller) with suitable methods in order to be able to analyze the response in the control loop and to design the controller according to the requirements. This description can be made in the time range (e.g. with differential equations) or in the frequency range (e.g. with a transfer function or a Bode diagram).

Many control-engineering transfer elements can be traced back to specific basic types or can be described by a linking of the same (Table 1).

The task of control-loop synthesis is to design, for a given controlled system, the matching controller (structure and parameters of the transfer element), which fulfills the above-mentioned requirements. There exists for this purpose a series of procedures (e.g. dynamic correction in the Bode diagram, root-locus curve procedures, pole specification, Riccati controller in the state space [4]), which are individually supplemented by specific functions or design steps.

A systematized procedure with the steps described in the following has proven to be useful in practice.

Designing a control task

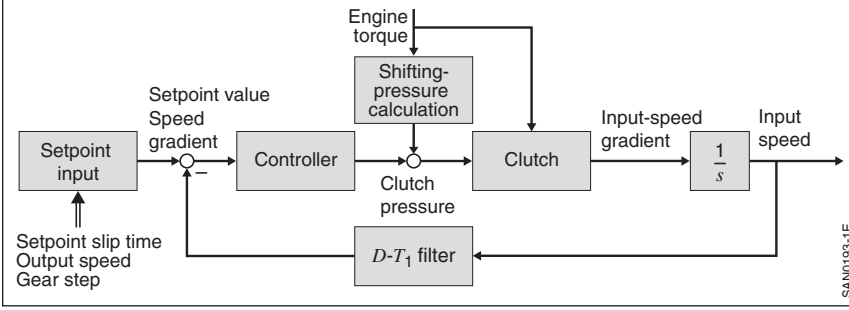
Control task

Typically the control task is not specifically formulated as such, but must be elaborated as the result of requirements in the specific technical process. This involves defining the open- and closed-loop control tasks in the system in order to settle the questions as to what is to be achieved with the control function and with which variables the objective is described. Controlled demand-response shifting in an automatic transmission is mentioned as an example. With this function, the clutch pressure of the shifting clutch is to be aligned during the gear shift to the speed gradient in such a way that the slip time remains constant under all operating conditions, even with variable parameters (e.g. friction coefficient).

System and block diagrams

It is helpful with these considerations, to draw up a system diagram in which the essential interaction between electronics, mechanics, hydraulics and pneumatics and all the sensors, actuators and bus systems can be clearly seen. From this, the control-engineering block diagram must be derived, from which in turn the functional interaction of all the open- and closed-loop-controlled functions with the system to be controlled can be seen (Figure 4). The functions are described in headline form, but not yet formulated in detail.

Figure 4: Control-engineering block diagram, for example of controlled demand-response shifting in an automatic transmission



It should be possible already to obtain a fundamental system understanding of the operative connections from the system diagram and the block diagram. As long as the system (mechanics, peripherals, hardware, etc.) is still in development it should at this point still be possible to exert an influence on the structural design of the system as contemplated by an overall mechatronic approach. The filling behavior of a hydraulically actuated clutch is mentioned as an example. This should be designed in such a way that a reproducible behavior with as little dead time as possible is provided on the basis of line cross-sections, volumes and sealing behavior.

Controlled system
The controlled system is then identified. This can be done theoretically (by modeling) or practically, e.g. by measuring the step response or the frequency response. It is recommended that both methods are followed and an adjustment carried out. Identification is a very comprehensive process, depending on the task. Sometimes it is sufficient to determine only the basic type and the order of the controlled system.

Controller design
The controller is designed on the basis of the result of the identification – this is the central task of controller synthesis. It is advisable for this first to be conducted theoretically using simulations, during which the controller parameters are set at the same time. If this step is sufficiently validated, startup follows on the real controlled system on the test bench or in the vehicle. Recursion steps are usually performed once again here in order to achieve further optimizations.

Design criteria
Over and above this fundamental sequence, reference is made to the following additional criteria:

Digital (discrete-time) control
The majority of control operations in a motor vehicle are implemented on microcontrollers. In this case, it is necessary to suitably establish the sampling time, oriented towards the dynamic response of the controlled system. It is necessary to ensure here that all the function algorithms can be calculated in the time available between two samplings.

Nonlinearities
In many cases, the simple linear methods described above are not sufficient, since real controlled systems contain nonlinearities (e.g. pressure-regulator characteristics, clutch characteristics, etc.). In simple cases, where static continuous nonlinearities are involved, these can be compensated for by an additional inverse-behavior transfer element. In the case of control operations with small signal amplitudes about an operating point, the system equations can be linearized at this point, otherwise more complex procedures are called for.

Structure switchovers
Many closed-loop control operations are first initiated by open-loop-controlled signals (e.g. first fill clutch, then set shifting pressure, then start shifting control). In these cases, it is necessary when switching from the open-loop- to the closed-loop-controlled state to ensure that this takes place smoothly and that the storage devices (integrators of the I elements) are correctly initialized.

Robustness
The control is typically designed with a "nominal" controlled system. In practice, however, manufacturing tolerances dictate that controlled systems are encountered which deviate up to a defined extent from this nominal definition. Furthermore, the parameters change over the service life, e.g. due to clutch wear or depending on third variables (temperature). In none of these cases should this give rise to significant functional impairments or instability of the control loop. Different procedures from the field of "robust control" or "adaptive control" are available to master these requirements.

Adaptive controllers

Motivation
Controlled systems often do not have a constant behavior. Parameters such as time constants and gains change in many cases. Even the structure of the system can change. Adaptations match open-and closed-loop control processes to a different system behavior. Examples include:

Manufacturing tolerances
Not all the products in a batch are 100 % identical. Because an individual adjustment is complicated, the system should adapt itself automatically to different parameters (e.g. adjustment data for automatic transmissions, [5]).

Wear
Parameters change in response to wear on a reproducible (e.g. increasing clutch control travel) or random level (e.g. friction coefficient of disks). Adaptations can compensate for this different behavior (e.g. clutch-control-travel adaptation for automated clutches, [6]).

Dependence on a third variable (e.g. temperature)
The viscosity of oil is highly temperature-dependent. Because these fluctuations will also occur in the short term (e.g. repeatedly every day), they must be compensated for (e.g. interference-torque observer, lockup-clutch control, [7], [8]).

Dependence on operating point of nonlinear systems
Nonlinear systems are frequently linearized at an operating point and then controlled by a linear controller (small-signal behavior). It is possible by means of an adaptation to take into account the different behavior at the operating points (e.g. adaptation of the shifting pressure for interlaced multiple gearshifts in an automatic transmission, [9]).

The various requirements to master this problem on an open-loop control level result in the desire for adaptive systems, which are described and defined in the following.

Definition of "adaptive control"
The behavior of the control is adapted to the changing properties of the controlled system and its signals. Adaptation procedures are thus essentially divided into two steps.
– Identification: Identification of the system behavior (system parameters) from time-variable values of the system.
– Adaptation: Adaptation of the open- or closed-loop control law as a reaction to a change of the system behavior.

Open-loop-controlled adaptations
Adaptation is performed by open-loop control, a forward-pointing structure. It is assumed here that the changing properties can be recorded by means of measurable external signals z (disturbances)

Figure 5: Open-loop-control adaptive system
y Controlled variable,
w Reference variable,
u Correcting variable,
z Disturbance.

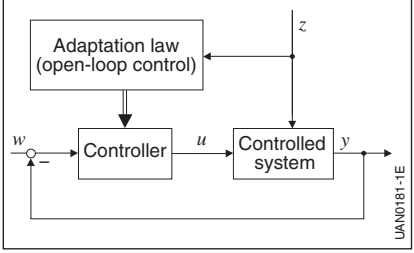
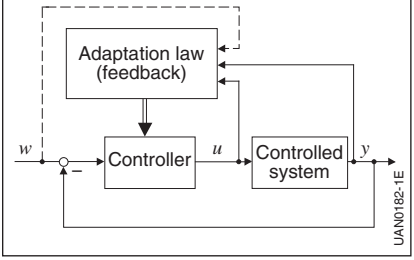


Figure 6: Adaptive system with feedback
y Controlled variable,
w Reference variable,
u Correcting variable.



and that the dependence of the control on these signals is known (Figure 5). There is no feedback on internal signals of the control loop to the setting of the controller.

Adaptation with feedback

In the case of an adaptation with feedback, the changing properties cannot be directly recorded and must be identified from measurable signals of the control loop. The identification can be a simple measurement or even a complex algorithm to estimate dynamic process models. In addition to the basic control loop, a second feedback loop is implemented by means of the adaptation law (Figure 6).

Essentially, an open-loop-controlled adaptation is initially recommended, i.e. known and metrologically recordable connections are used in open-loop-controlled form. The advantage of this forward structure can be compared with the advantages of open-loop control over closed-loop control. A feedback loop, which can possibly cause stability problems, is omitted. For the most part, open-loop-control adaptive systems are used in practical industrial applications.

Design notes

The following questions must be clarified when it comes to designing adaptive control:

– Which parameters and features must be adapted because they cannot be covered by a robust design?
– With which signals or variables can these parameters and features be determined?
– Does the system have to be specifically excited in order to determine these parameters and features, or can the adaptation take place during ongoing control operation?
– Can the adaptation be performed by open-loop control means or does feedback have to be provided for?
– How can the stability and convergence of adaptive control be tested?

A detailed treatment of the subjects areas of “Design of adaptive control operations” and “Identification of dynamic processes” can be found in [4], [10] and [11].

References

[1] DIN 19226: Control technology; terms and designations.
[2] W. Oppelt: Kleines Handbuch technischer Regelvorgänge. 5th Edition, Verlag Chemie GmbH, 1972.
[3] O. Föllinger: Regelungstechnik. 12th Edition, VDE Verlag GmbH, 2016.
[4] R. Isermann: Digitale Regelsysteme. 2nd Edition, Springer-Verlag, 2008.
[5] DE 102 007 040 485 A1, Abgleichdaten AT-Getriebe.
[6] DE 102 007 027 702 A1, Kupplungsweg-Adaption für automatisierte Kupplungen.
[7] DE 000 019 943 334 A1, Störmomentbeobachter, Wandlerkupplungsregelung.
[8] G. Bauer; C. Schwemer: Entwurf einer Wandlerkupplungsregelung unter Berücksichtigung nichtfunktionaler Anforderungen. AUTOREG 2008, 4th Symposium Baden-Baden, 12 and 13 February 2008, VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik.
[9] DE102 006 001899 A1, Adaption Schaltdruck bei verschachtelten Mehrfachschaltungen AT- Getriebe.
[10] R. Isermann: Identifikation dynamischer Systeme 1, 2. 2nd Ed., Springer-Verlag, 2012.
[11] R. Isermann: Mechatronische Systeme. 2nd Edition, Springer-Verlag, 2007.

Artificial intelligence

Introduction to AI

The aim of "artificial intelligence" (AI) is to understand how intelligent behavior works and how it can be implemented with information-technology means on computer systems. To this end attempts are made to enable machines, robots, and software systems to process and solve abstractly described tasks and problems independently without each step being programmed by humans. Important subfields are machine learning, knowledge representation, language understanding, and logical closure [1].

AI systems [2] are expected to have a certain degree of robustness, i.e., they should be able to adapt to their environment and changing conditions and learn [3].

The term system is furthermore based on the following definition: "A system is a set of interconnected components that has an expected behavior observed at the interface with its environment" [3].

History of AI

The establishment of AI as scientific discipline in its own right can be traced back to the "Dartmouth Summer Research Project on Artificial Intelligence", which took place in 1956 at Dartmouth College in Hanover (New Hampshire) and was attended by among others John McCarthy, Marvin Minsky, Herbert Simon, Alan Newell, and Claude Shannon [4].

AI as a science is based on two fundamental assumptions:

- 1.) Intelligence is founded on the ability of physical systems to represent regularities in the behavior of external phenomena in internal models, to analyze these representations, and to use the analysis results to control system behavior.
- 2.) The actual intelligence performance is the creation and manipulation of formal systems by formal systems.

Through the symbol-processing approach cognitive performances were taken as transformations of symbol structures with symbols being directly identified with meaningful units [5]. Some progress was made with this form of highly logic-oriented AI, but primarily in areas which could be efficiently described by logical formalisms.

This approach was less successful in fields characterized by high complexity, haziness or uncertainty. Emerging disillusionment led to a marked drop in research investment in AI, the so-called "AI winter".

Advances in AI

The increasing drop-off in prices with regard to computing capacity, memory, and connection costs and the associated emergence of cloud computing provided the material basis for implementing machine learning – specifically connectionist principles in the form of artificial neural networks (ANNs) – to a greater extent. Another often used term for this class of learning method is deep learning (see Machine learning). Here "knowledge" and "meaning" are implemented not by explicit symbolic representations but in activity patterns by means of "nodes" and "connections". The process is fundamentally different from classic algorithms not based on AI. In the case of classic algorithms the programmer uses instructions to describe how a given problem is to be solved. In the case of machine learning and ANNs the developer describes not the method but instead, by means of examples, the objective of the problem. In addition, they specifies a learning algorithm which independently searches for an approach towards a solution for the problem.

The resource requirements of elaborated ANNs must not be underestimated here. Thus, in one AlphaGo training cycle (see below) 5,000 tensor processing units were used over 40 days which required 192,000 KWh, the equivalent of the annual energy requirement of 32 households in Germany.

The dominant approaches in today's AI are those based on ANNs, where even greater importance could be attached in

future to the logic-based methods in the explanation of AI algorithms.

Significant milestones
Conceptional further developments of AI consistently resulted in spectacular breakthroughs which were observed with interest by the public at large:
– In 1996 “Deep Blue” beat the reigning world champion Gary Kasparov at chess.
– In 2005 “Stanley”, a modified VW Touareg, in the “DARPA Grand Challenge” succeeded in autonomously completing a course in the Mojave Desert of 212 km in less than seven hours and winning the race. 100 teams had already failed in this task the year before.
– In 2011 “Watson”, a text-understanding and semantic-search system, beat two human opponents in the quiz show “Jeopardy”.
– In 2012 “AlexNet”, a “deep” neural network, won the “ImageNet Large Scale Visual Recognition Challenge” by a long distance with a “super human performance”.
– In 2016 “AlphaGo” defeated the champion and best player Lee Sedol at the board game Go. This achievement was all the more remarkable since Go is significantly more complex than chess and it is impossible for supercomputers to search the entire solution space exhaustively.

These successes can lead to abstracting from specific ideal competition conditions and assuming that such systems can be easily transferred to open everyday situations. In actual fact, systems that exhibit impressive superhuman abilities under ideal conditions are must less robust against disruptions or influences that can arise in everyday situations.

Significant applications
Thanks to the significant algorithmic advances artificial intelligence has today already arrived on the scene in many products in our everyday lives.

Examples
Personal assistants
A prominent example is the personal assistant, with which users can interact for the most part via speech. Users can for example have questions about the current weather or the news answered simply by their mobile end device. For this purpose various AI algorithms run in the background
– to translate the audio signals into a representation of the question that is accessible to the computer,
– to find the information in a knowledge graph,
– to generate a voice output to communicate the information to the user.

Computer vision
A second application of AI that has in the meantime become widely used is computer vision (see Computer vision). The perception of the environment by means of images is for the most part a simple tasks for humans, but was for a long time an unsolved problem for machines. Today for the most part artificial neural networks (see Artificial neural networks) are used to draw conclusions from a camera image. Essentially, this features for example in driver-assistance systems, specifically the recognition of road signs, traffic lights, the lane, pedestrians, and other general road users. AI is similarly also used for monitoring with the assistance of cameras, for example to recognize exceptional situations in advance.

Significance of AI
These problems could not be technically solved without AI algorithms. A general rule of thumb for the applications of AI is: If the problem can be easily solved by classic methods, these are to be generally preferred. AI is exciting for applications in which the path to the solution cannot be described by a few simple rules as an algorithm.

Intelligent agents
AI systems are often based on the underlying concept of an “intelligent agent” which on account of its own knowledge appears in the position to act rationally in a given context and therefore is perceived

by external human observers to be intelligent. This concept of the attribution of intelligence by an external observer on account of actions that suggest intelligent knowledge processing was formulated by A. Newell in 1980 [7].

The fundamental characteristics of the generic architecture of an intelligent agent are the "Sense Plan Act" components and the interaction with its environment (Figure 1). From the environment the agent receives via its perception information, which it processes and implements through actions, which impact on the environment. Important prerequisites for successful interactions are a sufficiently stable and reliable perception on the part of the agent and the depiction of the perceived objects in a world or domain model which facilitates planning processes and – at least for the respective part of the world – sufficient semantic understanding. Added to this is the ability to act, i.e., turn what has been learned and the object of planning into reality.

The abilities of learning, representing the content and its further processing will be addressed in the following. To conclude, aspects of the security of AI systems will be addressed.

Machine learning

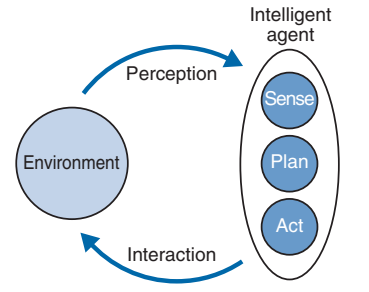
Learning as prediction

Learning is a process in which expertise is generated from experience, i.e.:

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks T, as measured by P, improves with experience E." [8]

Here a learner should successfully generalize from individual examples, i.e., inductively infer the general from the particular. Here the learner must infer from perceptible properties of an object the affiliation with a particular class. The learning objective is achieved if the learner was able to create an applicable hypothesis or a model which facilitates the correct prediction of the affiliation with a particular class on the basis of characteristics. Learning can thus be seen as a classification problem and the classification rules are encoded in the model [9]. Formally expressed: There must be a quantity $\mathbb{X}$ of objects which are to be classified. Furthermore, a quantity $\mathbb{Y}$ of possible annotations or labels with which the affiliation of examples with a class are identified. In the simplest case with an element from {0.1}. Moreover, a finite quantity $\mathbb{M}$ of annotated training examples from $\mathbb{X} \times \mathbb{Y}$, d.h. $\mathbb{M} = ((x_1,y_1)\ldots(x_m, y_m))$. The training examples derive from a

Figure 1: Generic architecture of an artificial intelligent agent
"Sense Plan Act" components;
Sense,
Plan,
Act.



probability distribution D from $\mathbb{X}$ and the annotations of all the examples were performed by an already existing classification function $f:\mathbb{X}\rightarrow\mathbb{Y}$, i.e., $y_i=f(x_i)$.

On the basis of these training data the learner formulates a hypothesis of a suitable classification model:

$$h:\mathbb{X}\rightarrow\mathbb{Y}.$$

The success of learning, i.e., the performance measure, can be determined by way of the prediction model generated by the learner. The error L of the prediction $h:\mathbb{X}\rightarrow\mathbb{Y}$ is therefore:

$$\begin{aligned} L_{D,f}(h) &= \mathbb{P}\,|(h(x)\neq f(x))| \\ &= D(\{h(x)\neq f(x)\}) \end{aligned}$$

The error thus relates to the probability distribution and the correct annotation function.

Example

Can a Bosch hammer or a Bosch orbital sander be seen in a photo? We could solve this task for example with machine learning. The quantity $\mathbb{X}$ is in this case photos in which either a Bosch hammer or an orbital sander can be seen. The quantity $\mathbb{Y}$ of possible annotations is here "Bosch hammer" and "orbital sander". Now we have in front of us a stack of photos, each with a "Post-it" note with the text (label) "Bosch hammer" or "orbital sander", depending on the content of the photo. This is the quantity $\mathbb{S}$, the example data, which we use to learn. The already existing classification algorithm was in this case a person who has assigned the correct label (from the quantity $\mathbb{Y}$) to the photos (from the quantity $\mathbb{X}$). With the aid of these learning examples we train the algorithm to solve the problem, given a new image we see a Bosch hammer or orbital sander.

The learning success can be easily measured here: we feed the algorithm a new image x_T. We obtain a prediction $h(x_T)$ of which of the tools is involved according to the learning algorithm. We compare this prediction with the probability, $f(x_T)$. Does this match, or is the learning algorithm wrong, is therefore $h(x_T)\neq f(x_T)$? If we repeat this test very frequently (infinitely often), we can estimate the probability that the algorithm is wrong in predictions. This probability is formally described by $\mathbb{P}\,|h(x_T)\neq f(x_T)|$ and produces a measure for the learning success. If this measure is one, $h(x)$ and $f(x)$ are always different, the algorithm never makes a correct prediction. If on the other hand the probability is zero, the learning was successful and the learning algorithm always produces a correct prediction.

Learning or training in the case of machine learning incidentally almost always means the following when put simply: Using data the learning process sets its parameters in such a way that there will be a correct prediction as often as possible.

Learning strategies

Supervised learning

The learning approach outlined above in which preclassified examples are used for learning is called "supervised learning". When the correctly annotated data are available, this is a highly efficient learning process. The effort involved in providing the data, annotating, and ensuring the correct annotation is however very high.

Unsupervised learning

A fundamentally different learning strategy is "unsupervised learning". Here objects are with regard to their similarity categorized in a target concept and appropriate clusters are created. In each case only one example – as typical as possible – is required as well as one metric (frequently Euclidean distance) and the objects are grouped and categorized in relation to their similarity by the target object.

Reinforcement learning

A further learning strategy is "reinforcement learning", in which an algorithm as with unsupervised learning receives non-annotated training examples and is to create a model. The modeling is to be steered in the desired direction by means of positive and negative feedback. This is almost a further development of the concept of evolutionary algorithms which modify themselves and in which the selection is controlled by means of fitness functions.

Important algorithms

Linear regression

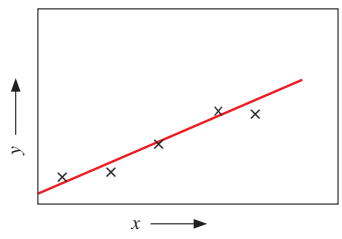
A very simple algorithm which however clearly illustrates machine learning is linear regression. Let us take for example some one-dimensional measured values, i.e., tuples at input variable x and measured value y.

In the case of linear regression we try to explain the measured values (the crosses in Figure 2) by a straight line $mx+c$. Quite specifically we must therefore define the two parameters m (straight-line gradient) and c (y axis intercept) in such a way that the distance to the measured values is minimal. We can achieve this for example with the aid of an optimization process. There are however also mathematical abbreviations so that this is not needed here and the optimal parameters can be determined directly. Optimization, i.e., adaptation of model parameters to data is called learning or training in the world of AI.

Artificial neural networks

"Artificial neural networks" (ANNs) constitute a much more complex process which is very widely used today. ANNs are based on a successful biological model – the human brain. In mathematical terms neural networks are the parallel and series connection of many simple modules.

Figure 2: Linear regression



Neuron

A module (neuron, Figure 3, [10]) assumes several input values ${}_xI$, … x_n, multiplies these values by weights w_1, … w_n, and applies a more often than not simple function to the sum total of these entries:

$$\varphi:\mathbb{R}\rightarrow\mathbb{R},\ \varphi\left(\sum_{i=1}^{n} w_i x_i\right).$$

This value forms the output value of the neuron.

Structure of a neural network

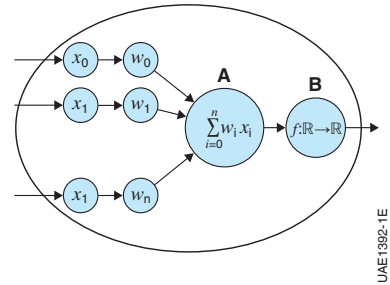
Numerous, in many cases millions of such operations are now connected in series (Figure 4). The output of a number of such parallel neurons forms the input of a further layer of neurons, thereby creating connections within a network. The last layer constitutes the output of the network. The size and structure of a network are dependent on the volume of training data and the complexity of the input values. Networks with up to 100 layers and over 100 million weights are used in image processing. A network however does not get "bigger" through learning.

Learning

ANNs usually engage in supervised learning, i.e., specifically the numerous weights of the network are adjusted by optimization in such a way that the ANN supplies the correct prediction for given

Figure 3: Artificial neuron

x Input values,
w Weights.
A Sum total of input value multiplied by weight,
B Activation function (e.g., $f(x) := \max(0.x)$).



examples. In this way the learning process is ended with the learned weights stored as the learning result.

This optimization process is very computation-intensive and was only made possible by new processor technologies, such as for example current-generation graphics cards.

Applications

Specifically, ANNs are used for example to classify objects, i.e., to estimate which category is involved for a given example. The last layer supplies a value as the output, for example 0 for Bosch hammer and 1 for orbital sander. Now many examples are presented to the network and the weights are adjusted in such a way that the prediction of the learning process is always as correct as possible (see section "Learning as prediction"). The entire learning process of the ANN is also called "deep learning".

So-called generative models constitute a second application aside from the classification of objects. Instead of predicting the class Bosch hammer or orbital sander in an image, the ANN is used to generate new data. Thus, the network could generate for example a new image, which constitutes a mixture of Bosch hammer and orbital sander. However, these tasks require much more complicated processes to adjust the weights of the ANN.

Representation of information

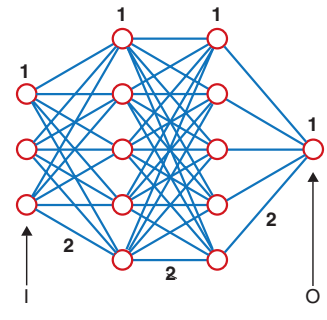
The representation of perceptions or experience content and subsequent further processing is one of the essential basic prerequisites for AI.

The manner of further processing is determinant for the appropriate form of representation. In those areas of AI which were greatly characterized by expert systems and the structuring of knowledge into easily communicable logical structures "knowledge representation" was practically synonymous with representation in logical formalisms together with an ontology, i.e., a systematic compilation of objects and their systematic correlation in a field of application which facilitated a semantic interpretation of the formulas.

Special neural networks, autoencoders, can be used to represent complex correlations. These are trained to learn an efficient representation for data and thus also important characteristics [11]. Significant application possibilities can be found in the compact representation of words with their context by "word embedding" in speech processing.

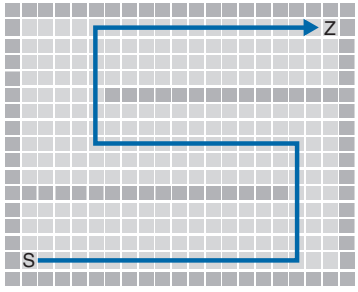
Generative adversarial networks (GANs) are based on a similar concept where the encoding and decoding are shared between two systems, a generator and a discriminator. The generator

Figure 4: Artificial neural network (ANN)
1 Neurons,
2 Connections.
I Input layer,
O Output layer.



has the task of creating a specific distribution of data. The discriminator attempts to distinguish these generated data from those of a "real" distribution which is present in the field of application. The aim of the generator is therefore to create such data which the discriminator evaluates as "real". GANs are used in a wide range of applications, from the creation of photo-realistic images to movement modeling in video sequences.

Figure 5: Simple example of a route-planning problem
S Start,
Z Destination.



Planning

Significance of planning

Behavior that seems rational or intelligent often requires a decision to be made or an action to be performed "now" in order to achieve an objective only very much later in the future. Many extreme challenging AI tasks have such a characteristic.

– In games like chess or Go a player must make a move in each round in order to win the game at some stage in the future. During each move they must estimate how the move will influence the game in the long term.
– Autonomous vehicles must plan on different levels. For instance, they must find a route to the destination. This is called route planning. However, the vehicles must also plan their motion, particularly the control commands such as steering angle or acceleration, in such a way that they reach their destination without collisions, even when other vehicles are moving in the same environment. This is called motion or behavior planning.

A prerequisite for planning is having an understanding of how an action will influence the future. The ability to plan also presupposes that an AI agent can answer the questions "What if …" or "What does the world look like after this action?"

More formally the agent requires a model of the environment. In a game the model is given by the game mechanism and the rules. In the navigation of an autonomous vehicle a model is obtained from the physics of the vehicle and the from the environment with all the obstacles.

Example of a planning problem

A simple and illustrative form of planning is the search for a route from a starting point to a destination (Figure 5). This planning problem can be calculated with simple graph-search methods. In this example the number of positions (also states) and possible movements are easily understandable. In real applications there can be extremely many states or action possibilities. In the game Go there are for example 10^{170} possible states on

the board. On a vehicle steering angle and acceleration can be continuously adjusted, which results in a practically infinite sphere of action. In the case of very large spheres of state and action, therefore, sampling-based methods are often used in the search in which actions are randomly selected and evaluated after a special statistical distribution.

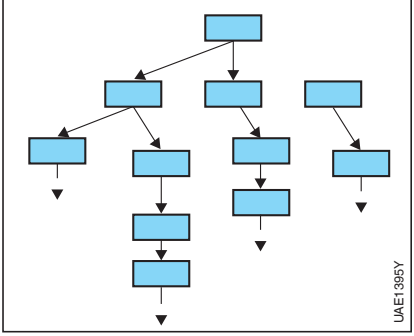
Problem when planning
Planning is made more difficult by the fact that the future for the most part cannot be predicted accurately. In games the future always depends on the opponent's moves. Autonomous vehicles must react to other road users. Unlike the case with route planning, a clear route leading to the destination cannot therefore be found. Rather a decision must depict and react to the "anticipated" future.

Computing time and computing capacity
A huge challenge in many planning problems is the required computing time and computing capacity. In a game like chess or Go a player (also a computer) only gets a limited time to think. In the case of a vehicle safety-critical decisions must be made in real time. The huge challenge is therefore to fashion the planning efficiently. An important concept here is the "leading" of the search algorithms. In this connection a function to be easily evaluated is often use to select the actions or states considered next by the program. These rules can be for example in a search on a geographical map the Euclidean distance to the destination (this is used with the so-called A* algorithm). But this selection function can also be more complex. The celebrated AlphaGo program [12], which in 2015 defeated the world's best human Go player, uses for example a very large ANN, to direct the search. This network was trained with countless simulated games. This trick enables AlphaGo, in a new game with a new (possibly previously unseen) game situation, nevertheless to develop a plan or a strategy very quickly and to make sensible moves.

New planning
Another trick to generate rational behavior is new planning. Instead of planning a complete route to the destination every time, a shorter plan is calculated. From this plan this first action is performed and a new planning is started. This process can be extremely efficient because on the one hand the computing time can be significantly reduced and on the other hand the system can also easily react to changes in the environment thanks to the continuous new planning. However, such an approach sets new challenges with regard to the completeness and optimality of the algorithms.

Intelligence
Planning is a classic and still very important sub-area of AI. The ability of an AI agent to plan and have strategic foresight is often, when viewed from the outside, perceived to be intelligence.

Figure 6: A search tree in which the nodes represent states and the arrows possible actions



IT security of AI systems

Comparison of an AI system with a classic program

An AI system differs from a classic computer program in that in the latter there is a marked separation between the application logic encoded in algorithms and the data used. The data flow is predefined by the structure of the algorithm. In AI systems the structure of a neural network for example is encoded in a specific algorithm. However, the weights of the network and consequently also the data flow through the system are dependent on the data. There is therefore a much closer correlation between data and system behavior than in conventional systems.

If a learned model is to be permanently saved, the structure and the values of the weights must therefore be recorded.

Manipulation possibilities

The software and the executable code of an AI system are subject to the same extent to all the risks and dangers of manipulation as any other software system that is run as an application on a computer system. In the IT environment security is usually considered from the perspectives of confidentiality, integrity, and availability. With regard to AI systems and intelligent agents system integrity plays a specific role because the intended functioning of a system is referenced with this term.

Manipulation of the control flow

"Hacking" systems has this very things as its object, namely the takeover and "external control" of a system through manipulation of the control flow. Instead of processing the original program commands, a system is made to execute any other code. In other words, the original algorithm is replaced by a different one. A hacker takes advantage of weaknesses in the system design, for example missing or insufficient input validation, thereby provokes errors in the system access, and "deposits" return addresses to code fragments which they have implanted.

Data poisoning

In an AI system there are furthermore additional possible ways to influence how the system works. Because adaptive systems are not programmed in the conventional sense but instead trained, system behavior can be altered by manipulating data from which the systems learn. This is called data poisoning.

In applications in which a system learns continuously there is a high degree of vulnerability to this type of attack. A system could alter its fundamental behavior in response to new data. For this reason, in the case of safety-critical requirements the weights of a network are "frozen" to prevent this. In addition, the perception of the systems can be disrupted by "adversarial attacks". This may result in a highly automated vehicle failing to recognize a traffic scene correctly and consequently not reacting appropriately.

Protective measures

While traditional computer systems could be protected extensively against attacks on their control flow – at least theoretically – with the reality behind the given possibilities remaining, this is not the case with AI systems. Full input validation cannot be provided in open application domains. It is therefore appropriate in AI systems, aside from observing the fundamental principles of secure system development in the system architecture, to provide further functionalities that actively observe, protect and stabilize system behavior.

This involves firstly the ability of a system to detect connected sequences of actions and to assign them to typical contexts and secondly the capacity to perceive itself in the time context and to compare current behavior with earlier behavior.

References

[1] S. Russell, P. Norvig: Artificial Intelligence: A Modern Approach. Prentice Hall, 3rd Edition, 2009.
[2] https://www.plattform-lernende-systeme.de/glossar.html
[3] “A system is a set of interconnected components that has an expected behavior observed at the interface with its environment”. See: J. Saltzer, F. Kaashoek: Principles of Computer System Design, Morgan Kaufman. Boston 2009, p. 8.
[4] G. Görz (Editor): Einführung in die Künstliche Intelligenz. Addison-Wesley. Bonn etc., 2nd Edition, 1995. p. 3.
[5] A. Newell, H. Simon: Computer Science as Empirical Inquiry: Symbols and Search. ACM, Vol. 19, No. 3. 1967.
[6] https://keplerlounge.com/artificial/intelligence/2019/03/24/alpha-go-zero.html
[7] A. Newell: The Knowledge Level. AI Magazine, Vol. 2, No. 2. 1981.
[8] T. Mitchell: Machine Learning. McGraw Hill. 1997.
[9] S. Sharlev-Shwartz, S. Ben-David: Understanding Machine Learning. Cambridge University Press, 2014.
[10] H. Wang, B. Raj: On the Origin of Deep Learning.
https://arxiv.org/abs/1702.07800, 2017.
[11] I. Goodfellow, Y. Bengio, A. Courville: Deep Learning. MIT Press. 2016. p. 493ff.
[12] D. Silver et al. “Mastering the game of Go with deep neural networks and tree search” Nature, 529, pp. 484–489 (2016).

Materials

Material parameters

Selecting material for a specific use is based on material parameters that describe the material properties. This process distinguishes between the physical material parameters, which result from the atomic structure, and the mechanical-technological material parameters, which are affected by manufacturing processes, for example.

Some physical material parameters are listed in tables 3 to 6 for a wide variety of materials.

Physical material parameters

Density

Density ρ is the ratio of the mass to the volume of a specific amount of substance (also see DIN 1306:1984-06 [1]). The unit of measurement is kg/m^3.

Melting temperature

The melting temperature denotes the temperature at which a material transforms from a solid to liquid state.

Boiling temperature

The boiling temperature denotes the temperature at which a material transforms from a liquid into a gas.

Heat of fusion

The specific heat of fusion of a solid (enthalpy of fusion) is the quantity of heat required to transform a material at fusion temperature from the solid to the liquid state. The unit of measurement is kJ/kg.

Heat of evaporation

The specific heat of evaporation of a liquid (enthalpy of evaporation) is the quantity of heat required to evaporate a liquid at constant boiling temperature. It is very pressure-sensitive. The unit of measurement is kJ/kg.

Thermal conductivity

Thermal conductivity is the quantity of heat that flows through a material sample with a defined surface and density due to temperature difference. In the case of liquids and gases (as opposed to solids), thermal conductivity is often highly dependent on temperature. The unit of measurement is W/(m·K).

Electric conductivity

Electric conductivity describes the physical property of a material to conduct electric current. According to the Wiedemann-Franz law, it is almost proportional to thermal conductivity. The unit of measurement is S/m (siemens per meter).

Coefficient of thermal expansion

The linear coefficient of thermal expansion α (coefficient of longitudinal expansion) indicates the relative change in length of a material during a temperature variation. For temperature change ΔT, the change in length is defined as $\Delta l = l\,\alpha\,\Delta T$. The unit of measurement is K^{-1}.

The cubic coefficient of thermal expansion (coefficient of volume expansion) is defined in a similar manner. The coefficient of volume expansion for gases is roughly 1/273 K^{-1}; For solids, it is roughly three times as large as the coefficient of linear expansion.

Heat capacity

The specific heat capacity (specific heat) is the quantity of heat required to raise the temperature of a substance. It is dependent on temperature.

In the case of gases, we distinguish between specific heat capacity at constant pressure and at constant volume (symbols c_p or c_v). This difference is usually negligible in the case of solid and liquid substances. The unit of measurement is kJ/(kg · K).

Permeability
Permeability μ describes the ratio of magnetic flux density B and magnetic field strength H (see magnetic field, relative permeability):

$$B = \mu H.$$

Permeability consists of the magnetic field constant μ_0 and the material-dependent relative permeability μ_r ($\mu_0 = 4\pi \cdot 10^{-7}$ A s/V m, $\mu_r = 1$ for vacuum). Depending on the application in which the magnetic material is used, there are roughly 15 types of permeability. These are defined according to modulation range and type of loading (direct-current or alternating-current field loading). The most important parameters are listed below.

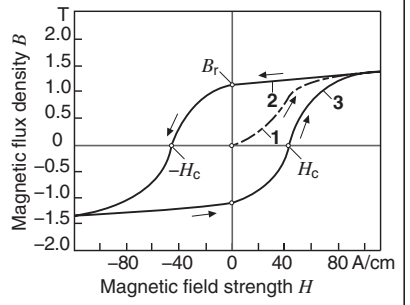
Initial permeability μ_a
The slope of the virgin curve (figure 1, also see hysteresis loop) for $H \rightarrow 0$ is designated as the initial permeability μ_a. In most cases, however, the slope for a specific field strength is specified (in mA/cm) rather than this limit value. Notation: μ_4 is the slope of the virgin curve for $H = 4$ mA/cm.

Maximum permeability μ_{max}
The maximum slope of the virgin curve is designated as the maximum permeability μ_{max}.

Recoil permeability μ_p
Recoil permeability μ_p (or μ_{rec}) is defined as the average slope of a retrograde magnetic hysteresis loop whose lowest point usually lies on the demagnetization curve:

$$\mu_p = \frac{\Delta B}{\Delta H \mu_0}.$$

Figure 1: Hysteresis loop of a magnetically hard iron grade
1 Rise path,
2, 3 Demagnetization curves.
H Magnetic field strength,
B Magnetic flux density,
B_r Remanence flux density,
H_c Coercive field strength.
The arrows indicate which magnetic flux density is set if the magnetic field strength is changed.



Temperature coefficient of magnetic polarization
The temperature coefficient of the magnetic polarization $TK(J_s)$ indicates the relative change in saturation polarization as temperature changes, it is given in percent per kelvin.

Temperature coefficient of coercive field strength
The temperature coefficient of the coercive field strength $TK(H_C)$ indicates the relative change in coercive field strength as temperature changes, it is given in percent per kelvin.

Curie point
The Curie point (Curie temperature T_C) is the temperature at which the magnetization of ferromagnetic and ferrimagnetic materials becomes zero and at which they behave like paramagnetic materials (see magnetic materials).

Mechanical-technological material parameters

Modulus of elasticity

The modulus of elasticity (E modulus) describes the linear relationship between stress and strain when deforming a solid body in the area of its elastic deformation (image 2). The unit of measurement is MPa. Examples of typical e-moduli are specified in table 1.

Table 1: Modulus of elasticity and Poisson's ratio for some materials

Material	Modulus of elasticity [MPa]	Poisson's ratio
Rubber	100	0.5
Fiber-reinforced plastic (e.g. PA66)	2,000	0.37
Aluminum	70,000	0.34
Titanium	110,000	0.28
Steel	200,000	0.3
Tungsten	400,000	0.28
Ceramics (e.g. Al_2O_3)	400,000	0.23

Poisson's ratio

The dimensionless Poisson's ratio is the proportionality factor between the longitudinal strain of a body under tension or compressive stress and the resulting transverse strain. Typical values are specified in table 1.

Yield point

The yield strength (yield point) is the amount of tensile stress starting at which a material experiences lasting plastic deformation. It is determined from the stress-strain curve (σ-ε curve) measured in the tensile test (see DIN EN ISO 6892-1:2009-12, [2]). The unit of measurement is MPa.

0.2% yield strength

The 0.2% yield strength ($R_{p0.2}$) is the amount of tensile stress that causes a lasting (plastic) strain of 0.2% in a material. The unit of measurement is MPa.

Tensile strength

The tensile strength R_m is the stress that is calculated in the tensile test from the maximum achieved tractive force relative to the output cross-section of a standardized material sample (figure 2 and figure 3). The unit of measurement is MPa.

Figure 2: Determining the modulus of elasticity

E modulus of elasticity,
R_m tensile strength,
$R_{p0,2}$ 0.2% yield strength,
A elongation at fracture,
A_g uniform strain.

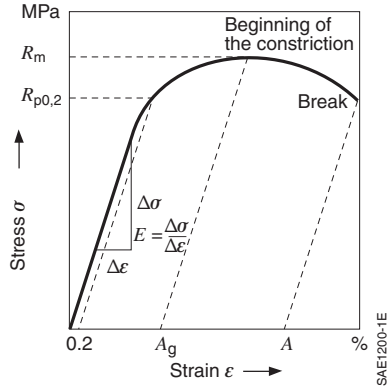
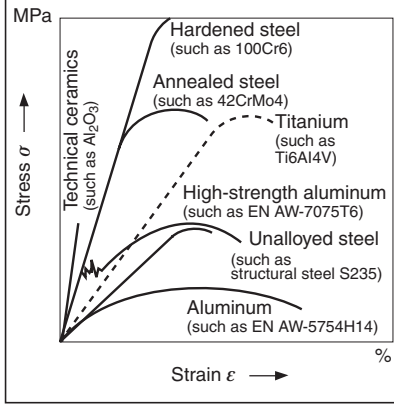


Figure 3: Stress-strain curve of some materials



Elongation at fracture
The elongation at fracture A is a dimension for the ductility (deformability) of a material. It indicates the lasting extension of a tensile specimen after the fracture relative to the initial length, specified in percent.

Fracture contraction
The fracture contraction Z, like the elongation at fracture, is a dimension for the ductility (deformability) of a material. It indicates in percent the change in cross-sectional area relative to the initial cross-section of a tensile specimen.

Uniform strain
The uniform strain A_g indicates the change in length in percent at which a tensile specimen goes through a plastic deformation without necking. For ductile materials, necking is caused on the specimen (cross-section reduction) after reaching the uniform strain (and tensile strength).

Vibration resistance
Vibration resistance is defined as the deformation and failure behavior of materials under cyclical stress. Typical vibration resistance values such as the fatigue limit have to be determined in often time-consuming Wöhler tests (fatigue tests named after August Wöhler) (image 4). These are usually carried out on electromechanical or electrohydraulic testing machines in the frequency range of 10 to 1,000 Hz.

A wide variety of test parameters influence the determined characteristic values (surface quality, test frequency, test duration, test medium, bend on tension or pressure). Here, the determined fatigue limit of a material is well below the structural strength (e.g. R_m). A simple estimate is possible using the FKM guideline (Forschungskuratorium Maschinenbau e.V., see Table 2). The tension-compression fatigue stress factor identifies the ratio of vibration resistance to tensile strength (tension-compression fatigue stress). For metals with a face-centered cubic crystal lattice, such as austenitic steel or aluminum, there is no fatigue limit since the stressability decreases when the stress duration increases.

Knowing the vibration resistance parameters is absolutely necessary for designing parts and components in mechanical engineering, since purely static stress is rare.

Fatigue strength under reversed bending stress
Fatigue strength under reversed bending stress identifies the stress in MPa that a material can withstand without breaking during cyclical bending stress (see vibration resistance).

Figure 4: Depicting the vibration resistance with the Wöhler curve with typical values for the number of stress cycles N
R_m tensile strength,
$S_{a,D}$ Fatigue limit.

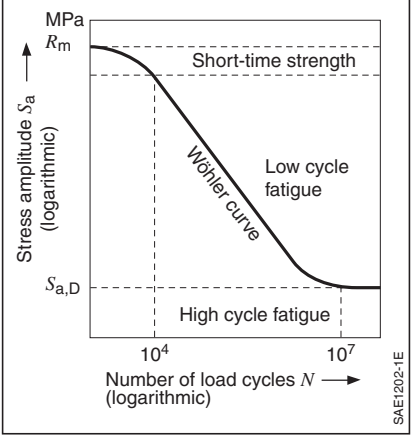


Table 2: Estimation of the vibration resistance

Material group	Tension-compression fatigue stress factor for the number of stress cycles $N = 10^6$
Aluminum, wrought and casting materials	0.30
Cast iron with flake graphite (GJL)	0.30
Cast iron with nodular graphite (GJS)	0.34
Stainless steel	0.40
Structural steel, tempering steel, etc.	0.45

Hardness
The hardness of a material identifies its resistance to the penetration of a body. Various methods for determining hardness are listed in the chapter "Heat treatment of metallic materials".

Fracture toughness
The fracture toughness is the resistance of a material to the spread of cracks. The critical stress intensity factor K_{Ic} is usually used as a characteristic of this. If K_{Ic} is known, the critical fracture load can be determined from crack length, or the critical crack length can be determined from the given external stress value.

Special characteristics for sintered material
Radial crushing strength
The radial crushing strength is a strength parameter which is specified in particular for sintered friction bearings. This is determined by buckling a hollow cylinder during the pressure test.

Porosity
The porosity is a dimensionless characteristic for sintered metals, particularly friction bearings. It describes the portion of part volume that consists of cavities and is not filled with actual material.

Compressive yield point
Comparable to the yield strength in the tensile test, the compressive yield point is added in the pressure test as a material characteristic (unit of measurement MPa). This identifies the stress starting at which an irreversible plastic deformation occurs in the material.

Special characteristics for springs
Bending stress
Bending stress occurs when stress is put on torsion springs. The maximum bending stress here should not exceed $0.7\,R_m$.

Shear stress
Shear stress occurs in the material when stress is put on pressure and tension springs. The maximum shear stress (maximum stress) should not exceed $0.5\,R_m$. Here, the shear stress is dependent on the coil diameter of the spring, spring force and wire diameter. The difference between maximum stress and minimum stress is also called stress range.

References for material parameters
[1] DIN 1306:1984-06: Density; concepts, presentation of values.
[2] DIN EN ISO 6892-1:2017-02: Metallic materials – Tensile testing – Part 1: Method of test at room temperature (ISO 6892-1:2016); German version EN ISO 6892-1:2016.
[3] D. Radaj, M. Vormwald: Ermüdungsfestigkeit – Grundlagen für Ingenieure [Fatigue strength – Principles for engineers]. 3rd Edition, Springer-Verlag, 2010.
[4] G. Gottstein: Physikalische Grundlagen der Materialkunde [Physical principles of material science]. 3rd Edition, Springer-Verlag, 2007.
[5] J. Schijve: Fatigue of Structures and Materials. 2nd Ed., Springer-Verlag, 2009.
[6] S. Suresh: Fatigue of Materials. 2nd Edition, Cambridge University Press, 1998.

Table 3: Properties of solids

Material/medium		Density	Melting point [1]	Boiling point [1]	Thermal conductivity [2]	Mean specific heat capacity [3]	Enthalpy of fusion ΔH [4]	Coefficient of longitudinal expansion [3]
		g/cm³	°C	°C	W/(m·K)	kJ/(kg·K)	kJ/kg	×10⁻⁶/K
Aluminum	Al	2.70	660	2,467	237	0.90	395	23.0
Aluminum alloys		2.60 to 2.85	480 to 655	–	70 to 240	–	–	21 to 24
Amber		1.0 to 1.1	< 300	Decomposes	–	–	–	–
Antimony	Sb	6.69	630.8	1,635	24.3	0.21	172	8.5
Arsenic	As	5.73	–	613 [5]	50.0	0.34	370	4.7
Asbestos		2.1 to 2.8	< 1,300	–	–	0.81	–	–
Asphalt		1.1 to 1.4	80 to 100	< 300	0.70	0.92	–	–
Barium	Ba	3.50	729	1,637	18.4	0.28	55.8	18.1 to 21.0
Barium chloride		3.86	963	1,560	–	0.38	108	–
Basalt		2.6 to 3.3	–	–	1.67	0.86	–	–
Beef tallow		0.9 to 0.97	40 to 50	< 350	–	0.87	–	–
Beryllium	Be	1.85	1,278	2,970	200	1.88	1,087	11.5
Bismuth	Bi	9.75	271	1,551	8.1	0.13	59	12.1
Bitumen		1.05	< 90	–	0.17	1.78	–	–
Boiler scale		< 2.5	< 1,200	–	0.12 to 2.3	0.80	–	–
Borax		1.72	740	–	–	1.00	–	–
Boron	B	2.34	2,027	3,802	27.0	1.30	2,053	5
Brass CuZn37		8.4	900	1,110	113	0.38	167	18.5
Brickwork		> 1.9	–	–	1.0	0.9	–	–
Bronze CuSn 6		8.8	910	2,300	64	0.37	–	17.5
Cadmium	Cd	8.65	321.1	765	96.8	0.23	54.4	29.8
Calcium	Ca	1.54	839	1,492	200	0.62	233	22
Calcium chloride		2.15	782	> 1,600	–	0.69	–	–
Cellulose acetate		1.3	–	–	0.26	1.47	—	100 to 160
Cement, set		2 to 2.2	–	–	0.9 to 1.2	1.13	–	–
Chalk		1.8 to 2.6	Decomposes into CaO and CO_2		0.92	0.84	–	–
Chamotte (fireclay)		1.7 to 2.4	< 2,000	–	1.4	0.80	–	–
Charcoal		0.3 to 0.5	–	–	0.084	1.0	–	–
Chromium	Cr	7.19	1,875	2,482	93.7	0.45	294	6.2
Chromium oxide	Cr_2O_3	5.21	2,435	4,000	0.42 [6]	0.75	–	–
Clay, dry		1.5 to 1.8	< 1,600	–	0.9 to 1.3	0.88	–	–
Cobalt	Co	8.9	1,495	2,956	69.1	0.44	268	12.4
Coke		1.6 to 1.9	–	–	0.18	0.83	–	–
Common salt		2.15	802	1,440	–	0.92	–	–
Concrete		1.8 to 2.2	–	–	< 1.0	0.88	–	–
Copper	Cu	8.96	1084.9	2,582	401	0.38	205	–
Cork		0.1 to 0.3	–	–	0.04 to 0.06	1.7 to 2.1	–	–
Corundum, fused		–	–	–	–	–	–	6.5 [7]
Cotton wadding		0.01	–	–	0.04	–	–	–

[1] At 1.013 bar. [2] At 20 °C. ΔH of chemical elements at 27 °C (300 K). [3] At 0 to 100 °C.
[4] At melting point and 1.013 bar. [5] Sublimated. [6] Powder form. [7] At 20 to 1,000 °C.

Table 3: Properties of solids (continued)

Material/medium		Density	Melting point [1]	Boiling point [1]	Thermal conductivity [2]	Mean specific heat capacity [3]	Enthalpy of fusion ΔH [4]	Coefficient of longitudinal expansion [3]
		g/cm³	°C	°C	W/(m·K)	kJ/(kg·K)	kJ/kg	×10⁻⁶/K
Diamond	C	3.5	3,820	–	–	0.52	–	1.1
Duromer								
Phenolic resin without filler		1.3	–	–	0.2	1.47	–	80
Phenolic resin with asbestos fibers		1.8	–	–	0.7	1.25	–	15 to 30
Phenolic resin with wood flour		1.4	–	–	0.35	1.47	–	30 to 50
Phenolic resin with fabric threads		1.4	–	–	0.35	1.47	–	15 to 30
Melamine resin with cellulose fiber		1.5	–	–	0.35	–	–	< 60
Foam rubber		0.06 to 0.25	–	–	0.04 to 0.06	–	–	–
Germanium	Ge	5.32	937	2,830	59.9	0.31	478	5.6
Glass (fused quartz)		–	–	–	–	–	–	0.5
Glass (window glass)		2.4 to 2.7	< 700	–	0.81	0.83	–	< 8
Gold	Au	19.32	1,064	2,967	317	0.13	64.5	14.2
Granite		2.7	–	–	3.49	0.83	–	–
Graphite, pure	C	2.24	< 3,800	< 4,200	168	0.71	–	2.7
Gray cast iron		7.25	1,200	2,500	58	0.50	125	10.5
Hard coal (anthracite)		1.35	–	–	0.24	1.02	–	–
Hard metal K 20		14.8	> 2,000	< 4,000	81.4	0.80	–	5 to 7
Hard rubber		1.2 to 1.5	–	–	0.16	1.42	–	50 to 90 [5]
Heat-conductive alloy NiCr 8020		8.3	1,400	2,350	14.6	0.50 [9]	–	–
HR foam, air-filled [8]		0.015 to 0.06	–	–	0.036 to 0.06	–	–	–
HR foam, freon-filled		0.015 to 0.06	–	–	0.02 to 0.03	–	–	–
Ice (0 °C)		0.92	0	100	2.33 [6]	2.09 [6]	333	51 [7]
Indium	In	7.29	156.6	2,006	81.6	0.24	28.4	33
Iodine	I	4.95	113.5	184	0.45	0.22	120.3	–
Iridium	Ir	22.55	2,447	4,547	147	0.13	137	6.4
Iron, pure	Fe	7.87	1,535	2,887	80.2	0.45	267	12.3
Lead	Pb	11.3	327.5	1,749	35.5	0.13	24.7	29.1
Lead oxide, litharge	PbO	9.3	880	1,480	–	0.22	–	–
Leather, dry		0.86 to 1	–	–	0.14 to 0.16	< 1.5	–	
Linoleum		1.2	–	–	0.19	–	–	–
Lithium	Li	0.534	180.5	1,317	84.7	3.3	663	56
Magnesium	Mg	1.74	648.8	1,100	156	1.02	372	26.1
Magnesium alloys		< 1.8	< 630	1,500	46 to 139	–	–	24.5
Manganese	Mn	7.47	1,244	2,100	7.82	0.48	362	22
Marble	$CaCO_3$	2.6 to 2.8	Decomposes into CaO and CO_2		2.8	0.84	–	–
Mica		2.6 to 2.9	Decomposes at 700 °C		0.35	0.87	–	3
Molybdenum	Mo	10.22	2,623	5,560	138	0.28	288	5.4
Monel metal		8.8	1,240 to 1,330	–	19.7	0.43	–	
Mortar, cement		1.6 to 1.8	–	–	1.40	–	–	–
Mortar, lime		1.6 to 1.8	–	–	0.87	–	–	–

[1] At 1.013 bar. [2] At 20 °C. [3] At 0 to 100 °C. [4] At melting point and 1.013 bar.
[5] At 20 to 50 °C. [6] At –20 to 0 °C. [7] At –20 to –1 °C.
[8] HR foam of phenolic resin, polystyrene, polyethylene or the like, values dependent on cell diameter and filler gas. [9] At 0 to 100 °C.

Table 3: Properties of solids (continued)

Material/medium		Density	Melting point [1]	Boiling point [1]	Thermal conductivity [2]	Mean specific heat capacity [3]	Enthalpy of fusion ΔH [4]	Coefficient of longitudinal expansion [3]
		g/cm³	°C	°C	W/(m·K)	kJ/(kg·K)	kJ/kg	×10⁻⁶/K
Nickel	Ni	8.90	1,455	2,782	90.7	0.46	300	13.3
Nickel silver	CuNi12Zn24	8.7	1,020	–	48	0.40	–	18
Niobium	Nb	8.58	2,477	4,540	53.7	0.26	293	7.1
Osmium	Os	22.57	3,045	5,027	87.6	0.13	154	4.3 to 6.8
Palladium	Pd	12.0	1,554	2,927	71.8	0.24	162	11.2
Paper		0.7 to 1.2	–	–	0.14	1.34	–	–
Paraffin		0.9	52	300	0.26	3.27	–	–
Peat dust, air-dried		0.19	–	–	0.081	–	–	–
Phosphorus (white)	P	1.82	44.1	280.4	–	0.79	20	–
Pitch		1.25	–	–	0.13	–	–	–
Plaster		2.3	1,200	–	0.45	1.09	–	–
Platinum	Pt	21.45	1,769	3,827	71.6	0.13	101	9
Plutonium	Pu	19.8	640	3,454	6.7	0.14	11	55
Polyamide		1.1	–	–	0.31	–	–	70 to 150
Polycarbonate		1.2	–	–	0.20	1.17	–	60 to 70
Polyethylene		0.94	–	–	0.41	2.1	–	200
Polystyrene		1.05	–	–	0.17	1.3	–	70
Polyvinyl chloride		1.4	–	–	0.16	–	–	70 to 150
Porcelain		2.3 to 2.5	< 1,600	–	1.6 [5]	1.2 [5]	–	4 to 5
Potassium	K	0.86	63.65	754	102.4	0.74	61.4	83
Quartz		2.1 to 2.5	1,480	2,230	9.9	0.80	–	8 [6]/14.6 [7]
Radium	Ra	5	700	1,630	18.6	0.12	32	20.2
Red bronze CuSn5ZnPb	Rb	8.8	950	2,300	38	0.67	–	–
Red lead, minium	Pb_3O_4	8.6 to 9.1	Forms PbO		0.70	0.092	–	–
Resin bonded fabric, paper		1.3 to 1.4	–	–	0.23	1.47	–	10 to 25 [8]
Resistance alloy	CuNi 44	8.9	1,280	< 2,400	22.6	0.41	–	15.2
Rhenium	Re	21.02	3,160	5,762	150	0.14	178	8.4
Rhodium	Rh	12.41	1,966	3,727	150	0.248	211	8.4
Roofing paper		1.1	–	–	0.19	–	–	–
Rosin		1.08	100 to 130	Decomposes	0.32	1.21	–	–
Rubber, raw		0.92	125	–	0.15	–	–	
Rubidium		1.53	38.9	688	58	0.33	26	90
Sand, quartz, dry		1.5 to 1.7	< 1,500	2,230	0.58	0.80	–	–
Sandstone		2 to 2.5	< 1,500	–	2.3	0.71	–	–
Selenium	Se	4.8	217	684.9	2.0	0.34	64.6	37
Silicon	Si	2.33	1,410	2,480	148	0.68	1,410	4.2
Silicon carbide		2.4	Decomposes above 3,000 °C		9 [9]	1.05 [9]	–	4.0
Sillimanite		2.4	1,820	–	151	1.0	–	–
Silver	Ag	10.5	961.9	2,195	429	0.24	104.7	19.2
Slag, blast furnace		2.5 to 3	1,300 to 1,400	–	0.14	0.84	–	–
Sodium	Na	0.97	97.81	883	141	1.24	115	70.6
Soft rubber		1.08	–	–	0.14 to 0.24	–	–	–

[1] At 1.013 bar. [2] At 20 °C. ΔH of chemical elements at 27 °C (300 K). [3] At 0 to 100 °C.
[4] At melting temperature and 1.013 bar. [5] At 0 to 100 °C. [6] Parallel to crystal axis.
[7] Perpendicular to the crystal axis. [8] At 20 to 50 °C. [9] At 1,000 °C.

Table 3: Properties of solids (continued)

Material/medium		Density	Melting point [1]	Boiling point [1]	Thermal conductivity [2]	Mean specific heat capacity [3]	Enthalpy of fusion ΔH [4]	Coefficient of longitudinal expansion [3]
		g/cm³	°C	°C	W/(m·K)	kJ/(kg·K)	kJ/kg	×10⁻⁶/K
Soot		1.7 to 1.8	–	–	0.07	0.84	–	–
Steatite		2.6 to 2.7	< 1,520	–	1.6 [6]	0.83	–	8 to 9 [5]
Steel, chromium steel		–	–	–	–	–	–	11
Steel, dynamo sheet		–	–	–	–	–	–	12
Steel, high-speed steel		–	–	–	–	–	–	11.5
Steel, magnet steel AlNiCo12/6		–	–	–	–	–	–	11.5
Steel, nickel steel 36 % Ni (invar)		–	–	–	–	–	–	1.5
Steel, sintered		–	–	–	–	–	–	11.5
Steel, stainless (18Cr, 8Ni)		7.9	1,450	–	14	0.51	–	16
Steel, tungsten steel (18 W)		8.7	1,450	–	26	0.42	–	–
Steel, unalloyed and low-alloy		7.9	1,460	2,500	48 to 58	0.49	205	11.5
Sulfur (α)	S	2.07	112.8	444.67	0.27	0.73	38	74
Sulfur (β)	S	1.96	119.0	–	–	–	–	–
Tantalum	Ta	16.65	2,996	5,487	57.5	0.14	174	6.6
Tellurium	Te	6.24	449.5	989.8	2.3	0.20	106	16.7
Thorium	Th	11.72	1,750	4,227	54	0.14	< 83	12.5
Tin (white)	Sn	7.28	231.97	2,270	65.7	0.23	61	21.2
Titanium	Ti	4.51	1,660	3,313	21.9	0.52	437	8.3
Tombac CuZn 20		8.65	1,000	< 1,300	159	0.38	–	–
Tungsten	W	19.25	3,422	5,727	174	0.13	191	4.6
Uranium	U	18.95	1,132.3	3,677	27.6	0.12	65	12.6
Vanadium	V	6.11	1,890	3,000	30.7	0.50	345	8.3
Vulcanized fiber		1.28	–	–	0.21	1.26	–	–
Wax		0.96	60	–	0.084	3.4	–	–
Wood [7]	Ash	0.72	–	–	0.16	2,1 to 2,9	–	in fiber direction 3 to 4, transverse to fiber 22 to 43
	Balsa	0.20	–	–	0.06		–	
	Beech	0.72	–	–	0.17		–	
	Birch	0.63	–	–	0.14		–	
	Maple	0.62	–	–	0.16		–	
	Oak	0.69	–	–	0.17		–	
	Pine	0.52	–	–	0.14		–	
	Poplar	0.50	–	–	0.12		–	
	Spruce, fir	0.45	–	–	0.14		–	
	Walnut	0.65	–	–	0.15		–	
Wood-wool building slabs		0.36 to 0.57	–	–	0.093	–	–	–
Zinc	Zn	7.14	419.58	907	116	0.38	102	25.0
Zirconium	Zr	6.51	1,852	4,377	22.7	0.28	252	5.8

[1] At 1.013 bar. [2] At 20 °C. ΔH of chemical elements at 27 °C (300 K). [3] At 0 to 100 °C.
[4] At melting temperature and 1.013 bar. [5] At 20 to 1,000 °C. [6] At 100 to 200 °C.
[7] Mean values for air-dried wood (humidity approximately 12 %).
Radial thermal conductivity; axial is approx. twice as high.

Table 4: Properties of liquids

Material/medium		Density[2]	Melting point[1]	Boiling point[1]	Thermal conductivity[2]	Specific heat capacity[2]	Melting enthalpy ΔH[3]	Evaporation enthalpy[4]	Coefficient of volume expansion
		g/cm³	°C	°C	W/(m·K)	kJ/(kg·K)	kJ/kg	kJ/kg	$\times 10^{-3}$/K
Acetone	$(CH_3)_2CO$	0.79	–95	56	0.16	2.21	98.0	523	–
Antifreeze-water mixture									
23% by vol.		1.03	–12	101	0.53	3.94	–	–	–
38% by vol.		1.04	–25	103	0.45	3.68	–	–	–
54% by vol.		1.06	–46	105	0.40	3.43	–	–	–
Benzene	C_6H_6	0.88	+5.5 [6]	80	0.15	1.70	127	394	1.25
Common-salt solution 20%		1.15	–18	109	0.58	3.43	–	–	–
Diesel fuel		0.81 to 0.85	–30	150 to 360	0.15	2.05	–	–	–
Ethanol	C_2H_5OH	0.79	–117	78.5	0.17	2.43	109	904	1.1
Ethanol 95% [9]		0.81	–114	78	0.17	2.43	–	–	–
Ethyl chloride	C_2H_5Cl	0.90	–136	12	0.11 [5]	1.54 [5]	69.0	437	–
Ethyl ether	$(C_2H_5)_2O$	0.71	–116	34.5	0.13	2.28	98.1	377	1.6
Ethylene glycol	$C_2H_4(OH)_2$	1.11	–12	198	0.25	2.40	–	–	–
Fuel oil EL		< 0.83	–10	> 175	0.14	2.07	–	–	–
Gasoline/petrol		0.72 to 0.75	–50 to –30	25 to 210	0.13	2.02	–	–	1.0
Glycerin	$C_3H_5(OH)_3$	1.26	+20	290	0.29	2.37	200	828	0.5
Hydrochloric acid 10%	HCl	1.05	–14	102	0.50	3.14	–	–	–
Kerosene		0.76 to 0.86	–70	> 150	0.13	2.16	–	–	1.0
Linseed oil		0.93	–15	316	0.17	1.88	–	–	–
Lubricating oil		0.91	–20	>300	0.13	2.09	–	–	–
Mercury [8]	Hg	13.55	–38.84	356.6	10	0.14	11.6	295	0.18
Methanol	CH_3OH	0.79	–98	65	0.20	2.51	99.2	1,109	–
Methyl chloride	CH_3Cl	0.99 [7]	–92	–24	0.16	1.38	–	406	–
m-xylene	$C_6H_4(CH_3)_2$	0.86	–48	139	–	–	–	339	–
Nitric acid, conc.	HNO_3	1.51	–41	84	0.26	1.72	–	–	–
Paraffin oil		–	–	–	–	–	–	–	0,764
Petroleum ether		0.66	–160	> 40	0.14	1.76	–	–	–
Rapeseed oil		0.91	± 0	300	0.17	1.97	–	–	–
Silicone oil		0.76 to 0.98	–	–	0.13	1.09	–	–	–
Sulfuric acid, conc.	H_2SO_4	1.83	+10.5 [6]	338	0.47	1.42	–	–	0.55

Table 4: Properties of liquids (continued)

Material/medium		Density[2] g/cm³	Melting point[1] °C	Boiling point[1] °C	Thermal conductivity[2] W/(m·K)	Specific heat capacity[2] kJ/(kg·K)	Melting enthalpy ΔH[3] kJ/kg	Evaporation enthalpy[4] kJ/kg	Coefficient of volume expansion ×10⁻³/K
Tar		1.2	–15	300	0.19	1.56	–	–	–
Toluene	C_7H_8	0.87	–93	111	0.14	1.67	74.4	364	–
Transformer oil		0.88	–30	170	0.13	1.88	–	–	–
Trichloroethylene	C_2HCl_3	1.46	–85	87	0.12	0.93	–	265	1.19
Turpentine oil		0.86	–10	160	0.11	1.80	–	293	1.0
Water		1.00 [10]	±0	100	0.60	4.18	332	2,256	0.18 [11]

[1] At 1.013 bar. [2] At 20 °C. [3] At melting temperature and 1.013 bar. [4] At boiling temperature and 1.013 bar. [5] At 0 °C.
[6] Setting temperature 0 °C. [7] At –24 °C. [8] For conversion from Torr to Pa, use 13.5951 g/cm³ (at 0 °C).
[9] Denatured ethanol. [10] At 4 °C. [11] Volume expansion on freezing: 9 %.

Table 5: Properties of water vapor

Absolute pressure bar	Boiling point °C	Evaporation enthalpy kJ/kg
0.1233	50	2,382
0.3855	75	2,321
1.0133	100	2,256
2.3216	125	2,187
4.760	150	2,113
8.925	175	2,031
15.55	200	1,941
25.5	225	1,837
39.78	250	1,716
59.49	275	1,573
85.92	300	1,403
120.5	325	1,189
165.4	350	892
221.1	374.2	0

Table 6: Properties of gases

Material/medium		Density [1]	Melting point [2]	Boiling point [2]	Thermal conductivity [3]	Specific heat capacity [3] kJ/(kg·K)			Enthalpy of evaporation [2]
		kg/m³	°C	°C	W/(m·K)	c_p	c_v	c_p/c_v	kJ/kg
Acetylene	C_2H_2	1.17	–84	–81	0.021	1.64	1.33	1.23	751
Air		1.293	–220	–191	0.026	1.005	0.716	1.40	209
Ammonia	NH_3	0.77	–78	–33	0.024	2.06	1.56	1.32	1,369
Argon	Ar	1.78	–189	–186	0.018	0.52	0.31	1.67	163
n-butane	C_4H_{10}	2.70	–138	–0.5	0.016	1.67	1.51	1.10	–
i-butane	C_4H_{10}	2.67	–145	–10.2	0.016	–	–	1.11	–
Blast-furnace gas		1.28	–210	–170	0.024	1.05	0.75	1.40	–
Carbon disulfide	CS_2	3.41	–112	–46	0.0073	0.67	0.56	1.19	–
Carbon dioxide	CO_2	1.98	–57 [4]	–78	0.016	0.82	0.63	1.30	368
Carbon monoxide	CO	1.25	–199	–191	0.025	1.05	0.75	1.40	–
Chlorine	Cl_2	3.21	–101	–35	0.009	0.48	0.37	1.30	288
City/town gas		0.56 to 0.61	–230	–2010	0.064	2.14	1.59	1.35	–
Cyanogen (dicyan)	$(CN)_2$	2.33	–34	–21	–	1.72	1.35	1.27	–
Dichlorodifluoromethane (= Freon F 12)	CCl_2F_2	5.51	–140	–30	0.010	0.61	0.54	1.14	–
Ethane	C_2H_6	1.36	–183	–89	0.021	1.66	1.36	1.22	522
Ethanol vapor	C_2H_4	2.04	–114	+78	0.015	–	–	1.13	–
Ethylene		1.26	–169	–104	0.020	1.47	1.18	1.24	516
Fluorine	F_2	1.70	–220	–188	0.025	0.82	0.61	1.35	172
Helium	He	0.18	–270	–269	0.15	5.20	3.15	1.65	20
Hydrogen	H_2	0.09	–258	–253	0.181	14.39	10.10	1.42	228
Hydrogen chloride	HCl	1.64	–114	–85	0.014	0.81	0.57	1.42	–
Hydrogen sulfide	H_2S	1.54	–86	–61	0.013 [1]	0.96	0.72	1.34	535
Krypton	Kr	3.73	–157	–153	0.0095	0.25	0.15	1.67	108
Methane	CH_4	0.72	–183	–164	0.033	2.19	1.68	1.30	557
Methyl chloride	CH_3Cl	2.31	–92	–24	–	0.74	0.57	1.29	446
Neon	Ne	0.90	–249	–246	0.049	1.03	0.62	1.67	86
Nitrogen	N_2	1.24	–210	–196	0.026	1.04	0.74	1.40	199
Oxygen	O_2	1.43	–218	–183	0.0267	0.92	0.65	1.41	213
Ozone	O_3	2.14	–251	–112	0.019	0.81	0.63	1.29	–
Propane	C_3H_8	2.00	–182	–42	0.018	1.70	1.50	1.13	–
Propylene	C_3H_6	1.91	–185	–47	0.017	1.47	1.28	1.15	468
Sulfur dioxide	SO_2	2.93	–73	–10	0.010	0.64	0.46	1.40	402
Sulfur hexafluoride	SF_6	6.16 [3]	–50.8	–63.9	0.011	0.66	–	–	117 [1]
Water vapor, 100 °C [5]		0.60	± 0	+100	0.025	2.01	1.52	1.32	–
Xenon	Xe	5.89	–112	–108	0.0057	0.16	0.096	1.67	96

[1] At 0 °C and 1.013 bar.
[2] At 1.013 bar.
[3] At 20 °C and 1.013 bar.
[4] At 5.3 bar.
[5] At saturation and 1.013 bar, also see Table 5.

Material groups

Material requirements

Technically and economically beneficial material use requires the knowledge of the application and the stress occurring in operation to derive the material requirements. The goal is to implement the desired function in the part or component and guarantee it throughout the planned period of use. The crucial loads here are usually of mechanical nature (strength, modulus of elasticity, see Table 1), but are often supplemented by corrosion, wear or temperature loads. Furthermore, specific physical properties (e.g. magnetism, conductivity) or a combination of those mentioned can be critical for material use.

Classification of materials

Various options exist for classifying the materials now used in technical applications. One example is subdividing into the four material groups:

– Metallic materials: smelted metals, sintered metals.
– Non-metallic inorganic materials: ceramics, glass.
– Non-metallic organic materials: natural substances, plastics.
– Composite materials.

Table 1: E-modulus of some materials (example)

Material group	Material type (selection)	Abbreviation Example	E-modulus [GPa]
Cast iron and malleable cast iron	Cast iron with lamellar graphite	EN-GJL-200	80 to 140
	Cast iron with spheroid graphite	EN-GJS-400	160 to 180
	White malleable cast iron	EN-GJMW-450	175 to 195
	Black malleable cast iron	EN-GJMB-450	175 to 195
Cast steel	Cast steel for general use	GE300	≈210
Steel	Unalloyed steels	C60	≈210
	Low-alloy steels	42CrMoS4	≈210
	Austenitic steels	X5CrNi18-10	≥190
	High-alloy tool steels	HS 6-5-4	≤230
Wrought copper alloys	High-conductivity copper	Cu-ETP	110 to 130
	Brass	CuZn30	115
	Nickel silver	CuNi18Zn20	135
	Tin bronze	CuSn6	100 to 120
Casting copper alloys	Tin cast bronze	CuSn10-C	100
	Red brass	CuSn7Zn4Pb7-C	100
Aluminum alloys	Wrought aluminum alloy	EN AW-AlSi1MgMn	65 to 75
	Casting aluminum alloy	EN AC-AlSi12Cu1	65 to 80
Magnesium alloys	Wrought magnesium alloy	MgAl6Zn	40 to 45
	Cast magnesium alloy	EN-MCMgAl9Zn1	40 to 45
Titanium alloys	Wrought titanium alloy	TiAl6V4	110
Tin alloys	Cast tin alloys for composite slide bearings	SnSb12Cu6Pb	30
Zinc alloys	Die-cast zinc alloy	GD-ZnAl4Cu1	70

Metallic materials

The metals generally have a crystalline structure. Their atoms are arranged in a regular crystal lattice. The valence electrons of the atoms are not bound to a specific atom, but are able to move freely within the metal lattice – a metallic bond is present. This special metal-lattice structure explains the characteristic properties of metals: the ductility and resulting high level of formability, the high electric conductivity, the high level of thermal conductivity, the low light transmission and the great optical reflective ability (metallic gloss).

The metallic materials used most often in technical applications include ferrous material. Iron base alloys are primarily used for this. An alloy is defined as a metallic material that consists of at least two chemical elements.

The ferrous materials are divided into groups for steels and cast iron materials. The main difference between the two groups is carbon content. The carbon content of steel is generally less than 2 % and over 2 % for cast iron (for details, see EN metallurgy standards).

Steels

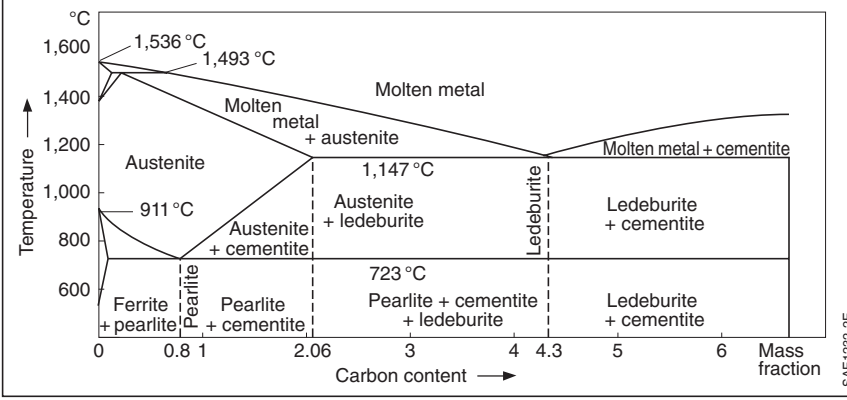
Structural compositions of steels

The majority of steels primarily consists of iron and other property-determining alloying elements such as chrome, nickel, vanadium, molybdenum and titanium. The most important alloying element is carbon, which is diffused in the iron lattice and present as iron carbide (Fe_3C), also called cementite. Various phases are present in the steel with their characteristic properties depending on the carbon content, the alloying elements and heat treatment (e.g. hardening, quenching and drawing). They can be found in the iron/carbon diagram (Figure 1). The most important phases are described below.

Ferrite

The ferritic structure exhibits a body-centered cubic atomic arrangement. It is soft, easily formable and ferromagnetic (e.g. S235). In unalloyed steels, ferrite is only present at very low carbon contents (< 0.02 %) as an individual phase (also called α phase).

Figure 1: Iron/carbon diagram



Pearlite
Pearlite has a lamellar iron-carbon structure, consisting of alternating associated areas of iron carbide (Fe_3C) and ferrite. This phase is present in steels with a carbon content higher than 0.02 %. As carbon content increase, the pearlite content present increases along with the ferrite until the structure consists entirely of pearlite at a carbon content of 0.8 % (e.g. C80). The pearlitic structure exhibits high strength.

Increasing the carbon content further results in the increased formation of iron carbide, which tends to separate out at the grain boundaries.

Austenite
If a steel is heated (between approximately 750 °C and 1,150 °C depending on the carbon content), the body-centered cubic iron lattice (α phase) transforms into a face-centered cubic iron lattice (γ phase). This austenitic structure can diffuse 100 times as much carbon as the ferritic structure. Ferrite and pearlite then form from the structure during the long cooling periods. In simple iron-carbon alloys, this lattice structure only exists at these high temperatures. Adding alloying elements such as nickel or cobalt allows the austenitic structure to be preserved stably even at room temperature. These types of austenitic alloys are very tough, easily formable, and resistant to heat and corrosion (e.g. X5CrNi1810).

Martensite
When cooling a steel very quickly from the austenite area during heat treatment (e.g. by quenching in water or oil), the carbon diffused in the lattice does not have enough time to transform into ferrite and pearlite. The austenitic structure abruptly deforms into a distorted lattice structure called martensite. The martensitic structure is very hard and brittle (e.g. 100Cr6).

Steel types
Along with many different organization and designation options, steels are often named based on their typical application, use or machining technique. Table 2 and Table 3 give an overview of properties and typical applications.

Structural steel
This is a simple steel that can usually be welded or machined easily. It is used in large quantities in general mechanical engineering and civil engineering and is reasonably priced (e.g. S235JR).

Free-cutting steel
This steel typically features high sulfur content, which causes short chips to form when machining (turning, milling). These chips can be extracted in the machinery without complications. These types of steels are usually used for volume components with low to medium-strength stresses (e.g. 11SMn30).

Tool steel
This is the umbrella term for a type of steel used for tools and molds. Depending on the application, it can feature high tensile strength, thermal stability, viscosity or corrosion resistance. This steel type can be classified as a cold-work steel (up to approx. 200 °C) or hot-work steel (up to approx. 400 °C), depending on the operating temperature. High-speed steels with high thermal stability (greater than 400 °C) and resistance to wear are also used (e.g. 90MnCrV8, X40CrMoV5-1, HS 6-5-2), particularly for machining tools such as drills and milling cutters.

Tempering steel
This type of steel is particularly well-suited for quench and draw heat treatment (hardening and draw tempering). This steel is often used for dynamically stressed shafts due to the high tensile strength and outstanding viscosity(e.g. 42CrMo4).

Table 2: Properties and typical applications for steels

Material group	Example materials	Surface hardness	Core hardness	Properties	Typical applications
Case-hardening steel (case-hardened and tempered)	C15	700 to 850 HV	200 to 450 HV	High resistance to wear	Moderately stressed gear wheels, joints
	16MnCr5	700 to 850 HV	300 to 450 HV	High resistance to wear	Piston pins, gear wheels
	18CrNiMo7-6	700 to 850 HV	400 to 550 HV	High resistance to wear, weldable	Gear wheels, drive parts
Nitriding steel (quenched, tempered and nitrided)	31CrMoV9	700 to 850 HV	250 to 400 HV	High resistance to wear	Highly-stressed piston pins, spindles
Ball and roller bearing steel (hardened and tempered)	100Cr6	60 to 64 HRC		High resistance to wear, high hardness and strength	Ball bearings, rolling elements, needle bearings
Cold-work steel (unalloyed, hardened and tempered)	C80	60 to 64 HRC		Surface hardened, tough core	Blades, chisels, striking tools
Cold-work steel (alloyed, hardened and tempered)	90MnCrV8	60 to 64 HRC		High dimensional stability, great edge retention and toughness.	Reamers, woodworking tools, thread cutters
	X210Cr12	60 to 64 HRC		Highly wear-resistant, great dimensional stability, high hardening properties	Cutting and punching tools, broaching tools, flange rollers, cold extrusion tools
Hot-work steel (hardened and tempered)	X40CrMoV5-1	43 to 45 HRC		High resistance to heat-related wear, great toughness and thermal conductivity	Forging dies, die-casting tools, extrusion dies
High-speed steel (hardened and tempered)	HS 6-5-2	61 to 65 HRC		High strength, great toughness, high hardness	Drills, countersinking tools
Stainless martensitic steel (hardened and tempered)	X20Cr13	≈40 HRC		Usable up to 550 °C, ferromagnetic, high resistance to wear	Pump parts, piston rods, nozzle needles, ship propellers
	X90CrMoV18	≥57 HRC		Good chemical resistance, resistance to wear and polishing properties	Cutting tools, ball bearings

Table 3: Properties and typical applications for steels

Material group	Example materials	Tensile strength [MPa]	Yield point [MPa]	Elongation at fracture A5 [%]	Fatigue strength under reversed bending stress [MPa]	Properties	Typical application examples
Hot-galvanized strip/ sheet	DX53D	≤ 380	≤ 260	≥ 30 (A80)	≥ 190	Good deep-drawing capability, hot-galvanized, higher strength	Automotive sheet metal parts e.g. engine hood, side panel
Structural steel	S235 JR	340 to 510	≥ 225	≥ 26	≥ 170	Good weldability, no preceding or retroactive heat handling	General construction parts, profiles
Free-cutting steel	11SMn30	380 to 570	–	–	≥ 190	High sulfur content for short chips, not suitable for welding	Moderately stressed volume components from the automotive industry such as shafts
	35S20	520 to 680	–	–	≥ 260	High sulfur content for short chips, not suitable for welding	Volume parts such as moderately stressed screws
Tempering steel	C45	700 to 850	≥ 490	≥ 14	≥ 280	Tough, heat treatable, moderate to high strength	Axles, bolts, screws
	42CrMo4	1,100 to 1,300	≥ 900	≥ 10	≥ 440		Crankshafts, axle shafts
	30CrNiMo8	1,250 to 1,450	≥ 1,050	≥ 9	≥ 500		Drive parts
Stainless ferritic steel (annealed)	X6Cr17	450 to 600	≥ 270	≥ 20	≥ 200	Good corrosion resistance, deep-draw capability, polishing properties, magnetic	Household devices, medical technology, sanitary
Stainless austenitic steel (solution annealed)	X5CrNi18 -10	500 to 700	≥ 190	≥ 45	–	Most frequently used corrosion resistant steel, good weldability, good deep-draw and polishing capability, non-magnetic	Chemical apparatus engineering, architecture, household objects
Cold strip (unalloyed)	DC05LC	270 to 330	≤ 180	≥ 40 (A80)	≥ 130	Good deep-draw properties	Automotive sheet metal parts e.g. engine hood, side panel

Table 4: Properties and typical applications for spring steel

Material	Diameter and thickness [mm]	Modulus of elasticity E [MPa]	Shear modulus G [MPa]	Minimum tensile strength [MPa]	Fracture contraction Z [%]	Permissible bending stress [MPa]	Permissible stress range for $N \geq 10^7$ [MPa]	Permissible maximum stress [MPa]	Properties and applications
Spring steel wire D, (patented and drawn cold-hammered)	1	206,000	81,500	2,230	40	1,590	380	1,115	Tension, compression, torsion springs, high static and low dynamic stress capacity
	3			1,840	40	1,280	360	920	
	10			1,350	30	930	320	675	
Stainless spring steel wire	1	185,000	73,500	2,000	40	1,400	–	1,000	Stainless springs
	3			1,600	40	1,130	–	800	
Heat-treated, alloyed valve-spring steel wire VD SiCr	1	200,000	79,000	2,060	50	–	430	1,030	High-strength spring material, high toughness e.g. passenger car valve springs
	3			1,920	50	–	430	960	
Spring steel strip Ck85	≤ 2.5	206,000	–	1,470	–	1,270	640	–	Highly stressed leaf springs
Stainless spring steel strip	≤ 1	185,000	–	1,370	–	1,230	590	–	Stainless leaf springs

Case-hardening steel
These steels feature a low carbon content (e.g. C15), which is usually increased up to 0.8 % by inserting the steel components into a carbonaceous furnace atmosphere (approx. 900 °C) in the edge region. The subsequent quenching causes martensite to form in the edge region and, in conjunction, high surface hardening at a simultaneously viscous core (ferritic, pearlitic). Applications include gear wheels and piston pins. Unlike nitriding, case hardening causes greater delays due to the martensitic hardening.

Nitriding steel
This steel forms a very hard and wear-resistant surface layer (approx. 10 µm) due to its chemical composition in combination with a corresponding heat treatment (nitriding at approx. 450 to 600 °C). Here, the viscous material core remains preserved, which makes this material particularly well-suited for gear wheels. There is less component deformation due to the lower temperature in comparison to case hardening (e.g. 31CrMoV9).

Spring steel
These are steels that have a particularly fine structure with high yield strength due to alloying silicon. They are especially well-suited for manufacturing springs in engine production (e.g. valve springs). Table 4 gives an overview of various spring steels.

Cast iron materials
Cast iron is defined as an iron-based alloy with a carbon content greater than approximately 2 %. The cast iron designation derives from these alloys' particularly high level of castability. One reason for this is the significantly lower melting temperature in comparison to steel (approximately 1,150 °C compared to approximately 1,500 °C).

Table 5 gives an overview of typical properties and applications for some materials.

Table 5: Properties and typical applications for cast irons

Material group	Example materials	Tensile strength [MPa]	Yield point [MPa]	Elongation at fracture A5 [%]	Fatigue strength under reversed bending stress [MPa]	Properties	Typical application examples
Cast iron with lamellar graphite (gray cast iron)	EN-GJL-200	≥ 200	–	–	90	Good castability, high vibration damping and heat capacity	Pump housing, cylinder head, machine bed, brake disc
Cast iron with spheroid graphite	EN-GJS-400-15	≥ 400	≥ 250	≥ 15	200	Higher strength and ductility than GJL	Crankshaft, crank-case, hubs
White malleable cast iron	EN-GJMW-400-5	≥ 400	≥ 220	≥ 5 (A3)	–	Good weldability, thin wall thicknesses, castable	Fittings
Black malleable cast iron	EN-GJMB-350-10	≥ 350	≥ 200	≥ 10 (A3)	–	Better machinability and hardening capability, can be manufactured with thick walls as white malleable cast iron	Fittings

Types of cast iron
The slow cooling during casting results in carbon mostly being present as graphite, typically in a ferritic-pearlitic matrix.

Gray cast iron
Depending on the formation of the graphite in layers or spheroids, the cast iron is referred to as either gray cast iron with lamellar graphite (GJL) or gray cast iron with spheroid graphite (GJS). The spheroid formation of the graphite is controlled in this process by treating the molten metal accordingly, such as with calcium. GJS features higher strength and better formability than GJL. Gray cast iron with vermicular graphite represents the middle ground between these two alloy types, both in terms of both graphite formation as well as properties.

Malleable cast iron
Along with these three types, the structure can also be affected by heat treatment (malleablizing) after casting. For malleable cast iron, the casted component made of white cast iron is subjected to an annealing process, where the iron carbide present in the matrix is converted into temper carbon. This increases the viscosity of the material. Based on the appearance of the fracture surface, these are referred to as white (annealed in an oxidizing, decarburizing atmosphere) and black malleable cast irons (annealed in a neutral atmosphere).

Cast steel
Cast steel is used for particularly highly stressed, weldable castings. In comparison to simple cast iron with lamellar graphite, castability is significantly worse while casting temperature and shrinkage are considerably higher.

Nonferrous metals
Along with the steels, some nonferrous metals (NF metals) have important technically relevant properties, such as those required in various applications. Special emphasis should be placed on copper alloys (heavy metal, density greater than 5 g/cm^3), which are used in many applications due to their high electric conductivity, and easily machinable, lightweight and nonmagnetic aluminum materials (light alloy, density less than 5 g/cm^3).

Classification of NF metals
NF metals (nonferrous metals and alloys with an iron percentage of less than 50 %) can generally be classified into wrought and cast alloys. Here, NF wrought alloys exhibit a very well malleable, ductile structure and can be brought to the desired shape by deep-drawing, bending, flanging, extruding or using similar deformation processes. Due to low ductility, shaping is limited to casting for NF casting alloys.

The properties of NF metals are frequently adapted to material requirements by adding alloying elements. Tables 6 through 9 give an overview of typical properties and applications of some materials.

Aluminum alloys
In aluminum alloys, adding alloying elements primarily aims to improve mechanical properties. For example, adding zinc or copper makes heat treatment possible, which results in high strength.

Copper alloys
Alloying zinc to copper forms brass (e.g. CuZn30), which leads to a higher strength compared to pure copper. Copper alloys with other alloying elements other than zinc are called bronzes. The most common bronzes are tin bronzes (e.g. CuSn6) with high strength and low ductility. Copper-nickel alloys (nickel-silver) are frequently used for plug contacts.

Table 6: Properties and typical applications for wrought copper alloys

Material group	Typical materials	Tensile strength [MPa]	Yield strength [MPa]	Fatigue strength under reversed bending stress (reference values) [MPa]	Properties	Application examples
High-conductivity copper	Cu-OF R220	220 to 260	≤140	70	High electric conductivity, solderable, weldable, deformable	Electronics, electrical engineering
Brass	CuZn30 R350	350 to 430	>170	110	Good strength properties, excellent cold-workability, can be soldered and brazed very well	Deep-drawn parts, fastening elements, screws, zippers
Nickel silver	CuNi18Zn20 R500	500 to 590	>410	–	Cold-workable, non-tarnishing, good spring properties	Relay springs, plugs
Tin bronze	CuSn6 R470	>470	≈ 350	190	Good cold-workability, good strength properties, good hardening properties, good wear and corrosion resistance, solderable	Springs, metal hoses, general machine and apparatus engineering

Table 7: Properties and typical applications for NF wrought alloys

Material group	Typical materials	Tensile strength [MPa]	Yield strength [MPa]	Fatigue strength under reversed bending stress (reference values) [MPa]	Properties	Application examples
Pure aluminum	EN AW-Al99,5 O	65	20	40	High conductivity, ductile, food-grade	Apparatus and tank engineering, deep-drawn parts
Self-hardening aluminum alloy	EN AW-AlMg2Mn0,8 H111	190	80	90	Seawater resistant, easily weldable	Automotive and ship engineering
Heat-treatable aluminum alloy	EN AW-AlSi1MgMn T6	310	260	90	Artificially aged, good combination of strength, corrosion resistance	Most frequently used heat-treatable aluminum alloy, profiles, vehicle frames, automobile transverse links
Heat-resistant aluminum alloy	EN AW-AlCu4MgSi (A) T4	390	245	120	Very good machinability, high temperature stability, good thermal stability	Components in hydraulic systems, aviation industry
High-strength aluminum alloy	EN AW-AlZn5,5MgCu T6	540	485	140	Good machinability, extremely high strength	Aviation industry, mechanical engineering
Universal aluminum casting alloy	EN AC-AlSi7Mg0,3 T6	290	210	80	Universal casting alloy, good mechanical properties, good corrosion resistance, very good weldability and machinability	Fittings, motor production, architecture
Eutectic casting alloy	EN AC-AlSi12Cu1 (Fe) DF	240	140	70	Excellent mold filling capability and casting characteristics, good chemical resistance	Thin-walled castings, fittings
Universal aluminum die-cast alloy	EN AC-AlSi9Cu3 (Fe) DF	240	140	70	Good processibility, low cost	Intake manifolds, transmission housings

Table 8: Properties and typical applications for NF casting alloys

Material group	Typical materials	Tensile strength [MPa]	Yield strength [MPa]	Fatigue strength under reversed bending stress (reference values) [MPa]	Properties	Application examples
High-strength casting alloy	EN AC-AlCu4Ti K T6	330	220	90	High strength and toughness	Transmission housings
Magnesium die-cast alloy	EN-MCMgAl9Zn1	200	140	50	Very light, low corrosion and thermal resistance, chips are combustible	Covers, housing, mobile telephone parts
Titanium forge alloy	TiAl6V4	890	820	–	High strength and corrosion resistance	Implants, aviation industry

Other NF metals
Along with aluminum and copper, the NF metals described below are of particular technical relevance.

Magnesium
At a density of approximately 1.75 g/cm^3, magnesium is significantly lighter than aluminum. The low melting temperature makes it particularly well-suited for castings. Typical applications in automotive engineering include car wheel rims, housing units and profiles.

Titanium
Hard-to-machine titanium alloys feature great corrosion resistance and a high level of thermal stability. Due to the favorable ratio of density and strength, these alloys lend themselves well to the aviation industry (e.g. air compressor turbine blades). Another application area is medical technology (implants, e.g. hip prosthetics) due to biocompatibility.

Sintered metals
Metal powder sintering
Sintered metals are generally manufactured using near-net-shape pressing of metal powders. This allows complicated shapes to be made from sintered metals at low costs, ready-to-install immediately or only requiring little finishing. After near-net-shape pressing the metal powder, a permanent connection between the grains is established by a running diffusion process at temperatures between 60 and 80 % of the melting temperature.

Metal injection molding
During MIM (metal injection molding), a process variant of sintering, a component is formed by injection molding metal powder and plastic compounds. After removing lubricants and binders with chemicals or heat, the molded articles retain their characteristic properties just like during metal powder sintering.

In addition to its chemical composition, the degree of porosity determines the properties and application of sintered metals to a large extent (see Table 10 and Table 11).

Table 9: Other alloys and cast copper alloys

Material group	Typical materials	Tensile strength [MPa]	Yield strength [MPa]	Fatigue strength under reversed bending stress (reference values) [MPa]	Properties	Application examples
Tin alloy	SnSb12Cu6Pb	–	60	28	Low hardness, good corrosion resistance	Sliding bearings
Die-cast zinc	ZP0410	330	250	80	Excellent casting characteristics, dimensional stability, surface quality	Thin-walled, low-stressed castings
Heat-conductor alloy	NiCr8020	650	–	–	High electrical resistance, high temperature stability	Heating filament
Resistance alloy	CuNi44	420	–	–	Low temperature coefficient, high oxidation stability	Resistors, potentiometer, heating filament
Cast tin bronze	CuSn10-C-GS	250	160	90	Good corrosion, resistance to wear	Fittings, pump housings
Red brass	CuSn7Zn4Pb7-C-GZ	260	150	80	Seawater resistant, good machinability, limp-home characteristics	Friction bearings, bushings

References for metallic materials

[1] DIN 30910-3: Sintered metal materials – Sintered-material specifications – Part 3: Materials for bearings and structural parts with bearing properties.
[2] DIN 30910-4: Sintered metal materials – Sintered-material specifications – Part 4: Materials for structural parts.

Table 10: Materials for bearings and structural parts with bearing properties[1]

		Permissible ranges				Representative examples					
Material	Material code	Density ρ	Chemical composition in % by mass	Radial breaking strength K'[2]	Hardness	Density ρ	Chemical composition in % by mass	Radial breaking strength K'[2]	Compressive yield point $\delta_{d\,0.2}$	Hardness	Thermal conductivity λ
	Sint-	g/cm³	Percent	N/mm²	HB	g/cm³	Percent	N/mm²	N/mm²	HB[2]	W/mK
Sintered iron	A 00	5.6 to 6.0	<0.3 C;<1.0 Cu;	>150	>25	5.9	<0.2 others; rest Fe	160	130	30	37
	B 00	6.0 to 6.4	<2 others; rest Fe	>180	>30	6.3		190	160	40	43
	C 00	6.4 to 6.8		>220	>40	6.7		230	180	50	48
Sintered steel containing Cu	A 10	5.6 to 6.0	<0.3 C; 1 to 5 Cu;	>160	>35	5.9	2.0 Cu;	170	150	40	36
	B 10	6.0 to 6.4	<2 others; rest Fe	>190	>40	6.3	<0.2 others; rest Fe	200	170	50	37
	C 10	6.4 to 6.8		>230	>55	6.7		240	200	65	42
Sintered steel containing Cu and C	B 11	6.0 to 6.4	0.4 to 1.5 C; 1 to 5 Cu; <2 others; rest Fe	>270	>70	6.3	0.6 C; 2.0 Cu; <0.2 others; rest Fe	280	160	80	28
Sintered steel containing higher Cu	A 20	5.8 to 6.2	<0.3 C; 15 to 25 Cu;	>180	>30	6.0	20 Cu;	200	140	40	41
	B 20	6.2 to 6.6	<2 others; rest Fe	>200	>45	6.4	<0.2 others; rest Fe	220	160	50	47
Sintered steel containing higher Cu and C	A 22	5.5 to 6.0	0.5 to 3.0 C;	>120	>20	5.7	2.0 C[3]; 20 Cu;	125	100	25	30
	B 22	6.0 to 6.5	15 to 25 Cu; <2 others; rest Fe	>140	>25	6.1	<0.2 others; rest Fe	145	120	30	37
Sintered bronze	A 50	6.4 to 6.8	<0.2 C; 9 to 11 Sn;	>120	>25	6.6	10 Sn;	140	100	30	27
	B 50	6.8 to 7.2	<2 others; rest Cu	>170	>30	7.0	<0.2 others; rest Cu	180	130	35	32
	C 50	7.2 to 7.7		>200	>35	7.4		210	160	45	37
Sintered bronze graphitic[4]	A 51	6.0 to 6.5	0.5 to 3.0 C;	>100	>20	6.3	1.5 C[4]; 10 Sn;	120	80	20	20
	B 51	6.5 to 7.0	9 to 11 Sn;	>150	>25	6.7	<0.2 others; rest Cu	155	100	30	26
	C 51	7.0 to 7.5	<2 others; rest Cu	>170	>30	7.1		175	120	35	32

[1] According to "Material Specification Sheets for Sintered Metals": DIN 30910-3, 2004 edition [1]. [2] Measured on calibrated bearings 10/16Ø·10.
[3] C is primarily present as free graphite. [4] C is present as free graphite.

Table 11: Sintered metals for structural parts[1]

		Permissible ranges			Representative examples						
Material	Material code	Density ρ	Chemical composition in % by mass	Hardness	Density ρ	Chemical composition in % by mass	Tensile strength R_m	Yield point $R_{p\,0.1}$	Elongation at break A	Hardness	E-module $E \cdot 10^3$
	Sint-	g/cm³	Percent	HB	g/cm³	Percent	N/mm²	N/mm²	Percent	HB	N/mm²
Sintered iron	C 00	6.4 to 6.8	<0.3 C; <1.0 Cu;	>35	6.6	<0.5 others; rest Fe	120	60	3	40	100
	D 00	6.8 to 7.2	<2 others; rest Fe	>45	6.9		170	80	8	50	130
	E 00	>7.2		>60	7.3		240	120	14	60	160
Sintered steel containing C	C 01	6.4 to 6.8	0.3 to 0.9 C; <1.0 Cu;	>70	6.6	0.5 C;	240	170	2	75	100
	D 01	6.8 to 7.2	<2 others; rest Fe	>90	6.9	<0.5 others; rest Fe	300	200	2	90	130
Sintered steel containing Cu	C 10	6.4 to 6.8	<0.3 C; 1 to 5 Cu;	>40	6.6	1.5 Cu;	200	140	2	55	100
	D 10	6.8 to 7.2	<2 others; rest Fe	>50	6.9	<0.5 others; rest Fe	250	180	3	80	130
	E 10	>7.2		>80	7.3		340	240	5	110	160
Sintered steel containing Cu and C	C 11	6.4 to 6.8	0.4 to 1.5 C; 1 to 5 Cu;	>80	6.6	0.6 C; 1.5 Cu;	390	290	1	115	100
	D 11	6.8 to 7.2	<2 others; rest Fe	>95	6.9	<0.5 others; rest Fe	460	370	2	130	130
	C 21	6.4 to 6.8	0.4 to 1.5 C; 5 to 10 Cu; <2 others; rest Fe	>105	6.6	0.8 C; 6 Cu; <0.5 others; rest Fe	470	360	<1	140	100
Sintered steel containing Cu, Ni and Mo	C 30	6.4 to 6.8	<0.3 C; 1 to 5 Cu;	>55	6.6	0.3 C; 1.5 Cu; 4.0 Ni;	360	290	2	100	100
	D 30	6.8 to 7.2	1 to 5 Ni; <0.6 Mo;	>60	6.9	0.5 Mo;	460	330	2	125	130
	E 30	>7.2	<2 others; rest Fe	>90	7.3	<0.5 others; rest Fe	570	390	4	160	160
Sintered steel Cu, Ni and Mo	C 31	6.4 to 6.8	<0.3 C; <3.0 Cu;	>50	6.6	0.2 C; 2.0 Ni;	320	220	1	100	100
	D 31	6.8 to 7.2	<5.0 Ni; 0.6 to 2 Mo;	>60	6.9	1.5 Mo;	380	260	2	120	130
	E 31	>7.2	<2 others; rest Fe	>90	7.3	<0.5 others; rest Fe	460	320	3	150	160
Sintered steel contains Mo and C	C 32	6.4 to 6.8	<0.3 to 0.9 C; <3.0 Cu;	>55	6.6	0.6 C; 2.0 Cu;	400	370	<1	140	100
	D 32	6.8 to 7.2	<5.0 Ni; 0.6 to 2 Mo; <2 others; rest Fe	>60	6.9	1.5 Mo; <0.5 others; rest Fe	520	480	1	180	130
Sintered steel containing P	C 35	6.4 to 6.8	<0.3 C; <1 Cu;	>70	6.6	0.45 P;	290	180	9	80	100
	D 35	6.8 to 7.2	0.3 to 0.6 P; <2 others; rest Fe	>80	6.9	<0.5 others; rest Fe	310	210	10	85	130
Sintered steel containing Cu and P	C 36	6.4 to 6.8	<0.3 C; 1 to 5 Cu;	>80	6.6	2.0 Cu; 0.45 P;	330	270	4	90	100
	D 36	6.8 to 7.2	0.3 to 0.6 P; <2 others; rest Fe	>90	6.9	<0.5 others; rest Fe	350	300	5	95	130

Table 11 (continued): Sintered metals for structural parts[1]

		Permissible ranges			Representative examples						
Material	Material code	Density ρ	Chemical composition in % by mass	Hardness	Density ρ	Chemical composition in % by mass	Tensile strength R_m	Yield point $R_{p\,0.1}$	Elongation at break A	Hardness	E-module $E \cdot 10^3$
	Sint-	g/cm³	Percent	HB	g/cm³	Percent	N/mm²	N/mm²	%	HB	N/mm²
Sintered steel containing Cu, Ni, Mo and C	C 39	6.4 to 6.8	0.3 to 0.9 C; 1 to 3 Cu;	>90	6.6	0.5 C; 1.5 Cu; 4.0 Ni;	480	350	1	140	100
	D 39	6.8 to 7.2	1 to 5 Ni; <0.6 Mo; <2 others; rest Fe	>120	6.9	0.5 Mo; <0.5 others; rest Fe	560	380	2	160	130
Stainless sintered steel AISI 316	C 40	6.4 to 6.8	<0.08 C; 10 to 14 Ni;	>95	6.6	0.06 C; 13 Ni;	330	250	1	110	100
	D 40	6.8 to 7.2	2 to 4 Mo; 16 to 19 Cr; <2 others; rest Fe	>125	6.9	2.5 Mo; 18 Cr; <0.5 others; rest Fe	400	320	2	135	130
AISI 430	C 42	6.4 to 6.8	<0.08 C; 16 to 19 Cr; <2 others; rest Fe	>140	6.6	0.06 C; 18 Cr; <0.5 others; rest Fe	420	330	1	170	100
AISI 410	C 43	6.4 to 6.8	<0.3 C; 11 to 13 Cr; <2 others; rest Fe	>165	6.6	0.2 C; 13 Cr; <0.5 others; rest Fe	510	370	1	180	100
Sintered bronze	C 50	7.2 to 7.7	9 to 11 Sn;	>35	7.4	10 Sn;	150	90	4	40	50
	D 50	7.7 to 8.1	<2 others; rest Cu	>45	7.9	<0.5 others; rest Cu	220	120	6	55	70
Sintered aluminum AlCuMgSi	E 73	2.55 to 2.65	4 to 6 Cu; <1 Mg; <1 Si; <2 other; rest Al.	>55	2.58 [2]	4.5 Cu; 0.6 Mg; 0.7 Si;	180	150 [4]	1	65	55 [5]
					2.58 [3]	<0.5 other; rest Al.	285	n.c.	<0.5	90	55 [5]
AlSiMgCu	F 75	2.60 to 2.66	2 to 3 Cu; <1 Mg; 13 to 16 Si; <2 other; rest Al.	>70	2.63 [2]	2.5 Cu; 0.5 Mg; 14 Si;	200	180 [4]	<0.5	90	78 [5]
					2.63 [3]	<0.5 other; rest Al.	300	n.c.	<0.5	125	78 [5]
AlZnMgCu	F 77	2.74 to 2.78	1.5 to 2.0 Cu; 2.2 to 2.8 Mg; 5.6 to 6.4 Zn; <2 other; rest Al.	>90	2.78 [2]	1.6 Cu; 2.6 Mg;	300	190 [4]	3	100	68 [5]
					2.78 [3]	6.0 Zn; <0.5 other; rest Al.	450	230 [4]	1.5	155	68 [5]

[1] According to "Material Specification Sheets for Sintered Metals": DIN 30910-4, 2010 edition [2].
[2] T1a sintered state and stored at room temperature for at least five days. [3] T6 solution annealed and artificially aged.
[4] Yield strength $R_{p0.2}$. [5] Determining the E-modulus using ultrasound.
n.c. not calculated.

EN metallurgy standards

Standardization of steels
(In accordance with DIN EN 10020, [1])
Steel is defined as an iron alloy, usually with a carbon content of ≤ 2%. Ferrous materials with a higher carbon content are usually classified as cast iron. Steels are classified in three classes, "unalloyed steels", "stainless steels" and "other alloyed steels".

Unalloyed steels
Unalloyed steels that do not reach the defined minimum alloying element content are then divided into unalloyed high-grade steels and unalloyed stainless steels. For unalloyed high-grade steels, defined requirements apply, such as those regarding toughness and malleability.

Unalloyed stainless steels feature a higher degree of purity and, therefore, improved properties such as high yield strength, hardenability, good toughness and weldability. They are usually intended for quenching and drawing or for surface hardening.

Stainless steels
Stainless steels feature a chrome content of at least 10.5 percent by mass and a carbon content of less than 1.2 percent by mass. They are subdivided further by nickel content (more or less than 2.5 percent by mass) and by the main properties, including corrosion resistance, heat resistance and temperature stability.

Other alloyed steels
These include steels with requirements in regards to aspects such as toughness, particle size or malleability. They are usually not intended for quenching and drawing or surface hardening. A distinction is made between alloyed high-grade steels and alloyed stainless steels.

Designation system for steels with material abbreviations
(In accordance with DIN EN 10027-1, [2])
The material abbreviations for steels are divided into two groups.
– Group 1: Material abbreviations that feature information on use and mechanical and physical properties.
– Group 2: Material abbreviations that contain information on chemical composition.

Abbreviated names of Group 1
These material abbreviations include notes on use and mechanical or physical properties. The prefix G indicates that it involves a cast steel material:

S, GS	For general steel construction,
P, GP	For pressure-vessel construction,
L	For pipeline construction,
E	Engineering steels,
B	Concrete reinforcing steels,
Y	Prestressing steels,
R	Rail steels,
H	Flat products made of higher-strength steels for cold forming,
D	Flat products for cold forming,
T	Packaging steel sheet and strip,
M	Electrical steel sheet and strip.

Examples:
S235JR

S	Generally for steel construction,
235	Yield strength in MPa,
JR	Notch impact work 27 J at 20 °C.

HC240LA

H	High-strength steel,
C	Cold-rolling,
240	Minimum yield strength in MPa,
LA	Low-alloyed.

Abbreviated names of Group 2
These material abbreviations contain references to the chemical composition. The prefix G indicates that this is a cast steel material:

C, GC	Unalloyed steels (Mn < 1%),
G	Unalloyed steels (Mn low-alloyed steels),
X, GX	High-alloyed steels,
PM	Powder metallurgy,
HS	High-speed steel.

Examples:
C85S

C	Unalloyed steel (< 1% Mn),
85	0.01 times the C content, i.e. 0.85 % C,
S	For springs.

42CrMo4

42	Low-alloyed steel (Mn ≥ 1%) 0.01 times the C content, i.e. 0.42 % C,
Cr	4 times the Cr content in percent, i.e. 1% Cr,
Mo	Non-specific percentage of alloyed constituents, Mo (< 1%).

X5CrNi18-10

X	High-alloyed steel,
5	0.05 % carbon,
18	18 % Cr,
10	10 % Ni.

HS 7-4-2-5

HS	High-speed steel Percentage of alloyed constituents in whole percentages in the order of tungsten – molybdenum – vanadium – cobalt:
7	7 % tungsten,
4	4 % molybdenum,
2	2 % vanadium,
5	5 % cobalt.

In addition, special requirements can be defined, such as for the type of coating or notes on the treatment condition. Example information could read as follows:

+H	With hardenability,
+CU	With copper coating,
+Z	Hot-dip galvanized,
+ZE	Electrolytically galvanized,
+C	Work-hardened,
+M	Thermomechanically shaped,
+Q	Quenched,
+U	Untreated.

Designation system for steels based on the numbering system
(In accordance with DIN EN 10027-2, [3])
In addition to their material abbreviation, all steels are defined by a material number in accordance with the following structure:
Major group + steel group + counter.

Material major group number:

0	Pig iron, ferro-alloys,
1	Steel,
2	Nonferrous heavy metals,
3	Light metals,
4 to 8	Non-metallic materials,
9	Unassigned for internal use.

Steel group number (selection):

00	Unalloyed ordinary low-carbon steels,
01 to 07	Unalloyed high-grade steels,
10 to 18	Unalloyed stainless steels,
40 to 49	Chemically stable alloying steels,
20 to 29	Alloying tool steels.

Example:
1.4301
(material abbreviation: X5CrNi18-10)

1	Steel,
43	Stainless steel, with more than 2.5 % Ni, without any Mo, Nb or Ti,
01	Count number.

Standardization of cast-iron materials
The structure and form of the carbon (carbide or graphite) and the graphite structure are important to the properties. In standardization, a distinction is made between the following four groups of cast iron:
– Lamellar graphite cast iron (DIN EN 1561, [4]),
– Malleable cast iron (DIN EN 1562, [5]),
– Spheroidal graphite cast iron (DIN EN 1563, [6]),
– Ausferritic spheroidal graphite cast iron (DIN EN 1564, [7]).

There are two systems for designating types of cast iron: In accordance with material numbers or abbreviations.

Designation system for cast iron with abbreviations
(In accordance with DIN EN 1560, [8])
The designation is alphanumeric and consists of up to six individual positions. The first position is EN (European standard). This is followed by G (cast) and J (iron). The character for denoting the graphite structure is in the third position:

L	Lamellar,
S	Spheroidal,
M	Temper carbon,
V	Vermicular,
N	Graphite-free, ledeburitic cast chilled iron,
Y	Special structure.

Notes on the microstructure and macrostructure can be placed at position 4 if necessary, such as (as an excerpt):

A	Austenite,
F	Ferrite,
Q	Quenched.

The fifth position contains information on tensile strength, impact energy with test temperature or for hardness. Alternatively, the fifth position can specify the chemical composition.

Finally, the sixth position contains additional requirements such as:

D	As-cast casting,
H	Heat-treated casting.

Examples:
EN-GJL-150

EN	European standard,
GJ	Cast iron,
L	Lamellar graphite,
150	Tensile strength in MPa.

EN-GJV-HV400

EN	European standard,
GJ	Cast iron,
V	Vermicular graphite,
HV400	Vickers hardness.

EN-GJS-SiMo30-8

EN	European standard,
GJ	Cast iron,
S	Spheroidal graphite,
Si	Silicon content 3 %,
Mo	Molybdenum content 0.8 %.

Designation system for cast iron with material numbers
(In accordance with DIN EN 1560, [8])
The designation consists of a total of six positions. The first and second positions contain "5." followed by information on the graphite structure at the third position:

1	Lamellar,
2	Vermicular,
3	Spheroidal,
4	Temper carbon.

The fourth position describes the matrix structure (selection):

1	Ferrite,
3	Pearlite,
5	Austenite.

The fourth position is followed by a double-digit count number (00-99) that identifies the individual material.

Nonferrous-metal alloys
As for ferrous materials, the EN standard also features two options for describing and identifying each nonferrous metals (NF) material and its alloys: The first designation system uses chemical symbols (abbreviations) and the second uses a numerical designation system.

Unlike steels, NF metals are each identified with their own numerical designation system. Thus, the different properties and requirements of, for example, aluminum, copper or zinc alloys are better taken into account.

Designation system for NF metals with abbreviation
The EN standard denotes nonferrous metals (NF metals) according to the following basic system:

Example:
EN AW–Al Si1MgMn T6

EN	European standard,

Code letter for metal:

A	Aluminum,
C	Copper,
M	Magnesium.

Code letter for processing:
W Wrought (wrought alloy),
C Casting (casting alloy).

Designation system with chemical symbols, e.g. AlSi1MgMn in this case:

Material condition, here, e.g.: T6.

The system lists the alloy metals in order of base metal with falling percentages.

Special numerical designation system for aluminum (Al) and aluminum alloys
For wrought Al products (Al and wrought Al alloys in accordance with DIN EN 573-1, [9]), the alloy is determined by four digits (Example 1); for Al casting alloys (in accordance with DIN EN 1706, [10]), by five digits (Example 2).

Example 1:
EN AW–6082 T6
EN European standard,
AW Wrought aluminum alloy,
6 Indicator 6 for alloy group (Al-Mg-Si alloys),
0 Original alloy (1, 2 for modifications),
82 Designation for the alloy with about 1% Si, 0.7% Mn and 0.9% Mg,
T6 Material condition T6 (solution-annealed and artificially aged).

Example 2:
EN AC -45200
EN European standard,
AC Casting aluminum alloy,
45 AlSi5Cu alloy group,
200 Number for individual alloy (here AlSi5Cu3Mn).

Code numbers for the alloy groups:
1 Pure aluminum,
2 With copper,
3 With manganese,
4 With silicon,
5 With magnesium,
6 With magnesium-silicon,
7 With zinc,
8 Other.

Material conditions (selection):
O Soft-annealed,
H Work-hardened,
H14 Work-hardened, 1/2 hard (for sheet metal),
T6 Solution-annealed and artificially aged.

References for EN metallurgy standards
[1] DIN EN 10020: Definition and classification of grades of steel.
[2] DIN EN 10027-1: Designation systems for steels – Part 1: Steel names.
[3] DIN EN 10027-2: Designation systems for steels – Part 2: Numerical system.
[4] DIN EN 1561: Founding – Grey cast irons.
[5] DIN EN 1562: Founding – Malleable cast irons.
[6] DIN EN 1563: Founding – Spheroidal graphite cast irons.
[7] DIN EN 1564: Founding – Ausferritic spheroidal graphite cast irons.
[8] DIN EN 1560: Founding – Designation system for cast iron – Material symbols and material numbers.
[9] DIN EN 573-1: Aluminium and aluminium alloys – Chemical composition and form of wrought products – Part 1: Numerical designation system.
[10] DIN EN 1706: Aluminium and aluminium alloys – Castings – Chemical composition and mechanical properties.

Magnetic materials

Materials which have ferromagnetic or ferrimagnetic properties are called magnetic materials and belong to one of two groups: metals (metals produced through melting or sintering) or nonmetallic inorganic materials. Composite materials such as soft magnetic composite materials and plastic-bonded permanent magnets are also playing an increasingly important role. They are characterized by their ability to exhibit a permanent external field or by their good magnetic flux conductivity (soft magnets).

In addition to ferromagnets and ferrimagnets, diamagnetic, paramagnetic, and antiferromagnetic materials also exist. They differ from each other in terms of their permeability μ, or the temperature dependence of their susceptibility κ. This variable gives the ratio of the magnetization of a substance to the magnetic field strength or excitation.

$$\mu_r = 1 + \kappa .$$

Classification

Ferromagnets and ferrimagnets

Both substances exhibit spontaneous magnetization which disappears at the Curie point (Curie temperature T_c). At temperatures above the Curie temperature, they behave like paramagnets. For susceptibility κ at $T > T_C$, the Curie-Weiss law applies as follows:

$$\kappa = \frac{C}{T - T_C},$$

where
C Curie constant,
T Temperature in K.

The saturation induction of ferromagnets is higher than for ferrimagnets because all magnetic moments are aligned in parallel. In the case of ferrimagnets, on the other hand, the magnetic moments of the two sublattices are aligned antiparallel to one another. Nevertheless, these materials are magnetic because the magnetic moments of the two sublattices have different magnitudes.

Diamagnets

For diamagnets, susceptibility κ_{Dia} is independent of the temperature.

Paramagnets

For paramagnets, susceptibility κ_{Para} drops as temperature increases. In this case, the Curie law states:

$$\kappa_{Para} = \frac{C}{T} .$$

Antiferromagnets

As in the case of ferrimagnets, adjacent magnetic moments are aligned antiparallel with respect to one another. As they are of equal magnitude, the effective magnetization of the material is zero.

At temperatures above the Néel point (Néel temperature T_N), antiferromagnets behave like paramagnets. At $T > T_N$, the susceptibility is:

$$\kappa = \frac{C}{T + \Theta},$$

where
Θ Asymptotic Curie temperature.

Examples of antiferromagnets: MnO, MnS, $FeCl_2$, FeO, NiO, Cr, V_2O_3, V_2O_4.

Soft magnetic materials

The figures specified in Table 1 are from the applicable DIN standards (soft-magnetic metallic materials, DIN IEC 60404-8-6, [1]). Many material qualities defined in this standard relate to the materials in DIN 17405 (DC relays, [2]) and DIN-IEC 60740-2 (transformers and reactors, [3]).

Designation (composition)

"Code letter" "Number 1" "Number 2" – "Number 3".

The code letter specifies the alloy class: "A" pure iron, "C" silicon-iron (SiFe), "E" nickel-iron (NiFe), "F" cobalt-iron (CoFe).

Number 1 indicates the concentration of the main alloy element.

Number 2 defines the different curves: 1 indicates a round hysteresis loop, 2 indicates a rectangular hysteresis loop.

Table 1: Soft-magnetic metallic materials

Magnet type	Alloying constituents by mass	Static magnetic properties											AC test data, 50 Hz [1]		
		Coercive field strength $H_{c(max)}$ in A/m Thickness in mm		Minimum magnetic polarization in tesla (T) at field strength H in A/m									Measuring point $\hat{H}$ in A/m	Minimum amplitude permeability μ_r Sheet thickness in mm	
	%	0.4 to 1.5	>1.5	20	50	100	300	500	800	1,600	4,000	8,000		0.30 to 0.38	0.15 to 0.20
A – 240	100 Fe	240	240				1.15	1.30			1.60		Not suitable for AC applications.		
A – 120	100 Fe	120	120				1.15	1.30			1.60				
A – 60	100 Fe	60	60				1.25	1.35			1.60				
A – 12	100 Fe	12	12			1.15	1.30	1.40			1.60				
C1 – 48	0 to 5 Si (typical 2 to 4.5)	48	48			0.60	1.10	1.20			1.50				
C1 – 12	0 to 5 Si (typical 2 to 4.5)	12	12			1.20	1.30	1.35			1.50				
C21 – 09	0.4 to 5 Si (typical 2 to 4.5)												1.60	900	750
C22 – 13	0.4 to 5 Si (typical 2 to 4.5)												1.60	1,300	–
E11 – 60	72 to 83 Ni	2	4	0.50	0.65	0.70		0.73			0.75		0.40	40,000	40,000
E21	54 to 68 Ni	Not suitable for this thickness												On agreement	
E31 – 06	45 to 50 Ni	10	10	0.50	0.90	1.10		1.35			1.45		0.40	6,000	6,000
E32	45 to 50 Ni	Not suitable for this thickness												On agreement	
E41 – 03	35 to 40 Ni	24	24	0.20	0.45	0.70		1.00			1.18		1.60	2,900	2,900
F11 – 240	47 to 50 Co		240				1.40		1.70	1.90	2.06	2.15	As agreed between manufacturer and buyer.		
F11 – 60	47 to 50 Co	60					1.80		2.10	2.20	2.25	2.25			
F21	35 Co	300							1.50	1.60	2.00	2.20			
F31	23 to 27 Co	300									1.85	2.00			

[1] Data applies to laminated rings.

The significance of Number 3 following the hyphen varies according to the individual alloy. It indicates the minimum initial permeability μ_a/1,000 in nickel alloys; with other alloys, it designates the maximum coercive field strength in A/m. The properties of these materials are strongly geometry-dependent and highly application-specific. The material data quoted in extracts from the standard can therefore provide only a very general overview of the properties of these materials.

Electrical steel sheet and electrical steel strip

Applications and properties

Packages used for transformers or as stator or rotor packages for electric motors are composed of electrical steel strips, typically by packaging individual sheet metal plates. Electrical steel strips are often provided as long strips (wide strips or split strips) in a wound shape (called coils), but they are also available as panels (electrical steel sheet).

Electrical steel strips typically consist of iron-silicon alloys ([4] and [5]); other materials such as aluminum and manganese may function as alloy constituents in small quantities. Densities between 7.65 and 7.85 g/cm³ result depending on the composition. They are characterized by low remagnetization loss, high polarization and high permeability compared to other soft magnetic ferrous materials. The static coercive field strength H_c is typically 100 to 300 A/m for A and K types (see Designations), the maximum permeability μ_{max} is at about 5,000. For S and P types, the values are in the order of magnitude of 1 A/m for H_c and about 30,000 for μ_{max}.

Grain oriented electrical steel strip

Both grain oriented and non-oriented electrical steel strips exist. Grain oriented electrical steel strips feature a crystal structure (cubic structure), meaning the grains are oriented in a preferred direction by rolling and annealing steps during manufacturing. This gives the strip material a preferred magnetic direction in the strip direction. The magnetic properties depend on the direction (anisotropic magnetic properties). Grain oriented electrical steel strip is primarily used in components where exceptionally good magnetic flux in a specific direction comes into play (e.g. transformers, choke coils and converters).

Non-oriented electrical steel strip

In non-oriented electrical steel strip, however, there is not a structure. In other words, the crystallographic orientation of the grains is almost randomly arranged. The magnetic properties do not depend on the direction (isotropic magnetic properties). For this reason, non-oriented electrical steel strip is used in components where magnetic flux is not limited to a preferred direction (e.g. electric motors and alternators) [6].

Non-oriented electrical steel strips are usually delivered in a finally annealed (fully-finished) state. In other words, final annealing for optimizing the magnetic properties takes place at the manufacturer of the electrical steel strips. For electrical steel strips that are not finally annealed (semi-finished), final annealing still has to be carried out. This is typically done to the finished laminated core.

Coating

Electrical isolation of the individual layers suppresses eddy currents for alternating-field magnetization, reducing remagnetization losses.

Identifications

The designations for electrical steel sheets and electrical steel strips (ESS) is defined in DIN EN 10027-1 [7]:

Code letter 1 Number 1 – Number 2 Code letter 2 (see below for an example).

The first code letter is “M” for all varieties. Number 1 is one hundredfold the maximum magnetic reversal loss at 50 Hz and 1.5 T (types A and K) or 1.7 T (types S and P)in W/kg. Number 2 is one hundredfold the product’s nominal depth in mm.

Table 2: Properties of electrical steel strips and electrical sheet steel

Sheet grade		Nominal thickness	Maximum remagnetization loss in W/kg at 50 Hz and activation of			Minimum magnetic polarization in tesla (T) in an AC field at field strength H in A/m			Applications
Abbreviated name	Material number	mm	1.0 T	1.5 T	1.7 T	2,500	5,000	10,000	
M270-35A	1.0801	0.35	1.10	2.7	–	1.49	1.60	1.70	
M330-35A	1.0804	0.35	1.30	3.3	–	1.49	1.60	1.70	
M330-50A	1.0809	0.50	1.35	3.3	–	1.49	1.60	1.70	
M530-50A	1.0813	0.50	2.30	5.3	–	1.56	1.65	1.75	
M800-50A	1.0816	0.50	3.60	8.0	–	1.60	1.70	1.78	Electric motors
M400–65A	1.0821	0.65	1.70	4.0	–	1.52	1.62	1.72	
M1000–65A	1.0829	0.65	4.40	10.0	–	1.61	1.71	1.80	
M800–100A	1.0895	1.00	3.60	8.0	–	1.56	1.66	1.75	
M1300–100A	1.0897	1.00	5.80	13.0	–	1.60	1.70	1.78	
M340-50K	1.0841	0.50	1.42	3.4	–	1.54	1.62	1.72	
M560-50K	1.0844	0.50	2.42	5.6	–	1.58	1.66	1.76	
M660-50K	1.0361	0.50	2.80	6.6	–	1.62	1.70	1.79	Low-power motors for industrial and household appliances (e.g. washing machine motors, microwave transformers, refrigerator compressors)
M1050-50K	1.0363	0.50	4.30	10.5	–	1.57	1.65	1.77	
M390-65K	1.0846	0.65	1.62	3.9	–	1.54	1.62	1.72	
M630-65K	1.0849	0.65	2.72	6.3	–	1.58	1.66	1.76	
M800-65K	1.0364	0.65	3.30	8.0	–	1.62	1.70	1.79	
M1200-65K	1.0366	0.65	5.00	12.0	–	1.57	1.65	1.77	
						At field strength H = 800 A/m			
M140-30S	1.0862	0.30	–	0.92	1.40	1.78			Converters, transformers, choke coils
M150-30S	1.0861	0.30	–	0.97	1.50	1.75			
M111-30P	1.0881	0.30	–	–	1.11	1.88			

Code letter 2 provides type data:
- "A" cold rolled non-oriented electrical steel sheet and strip delivered in the fully processed state (DIN EN 10106, [8]).
- "S" conventional grain-oriented electrical steel sheet or "P" grain-oriented electrical steel strip with high permeability, both types are delivered in a fully processed state (DIN EN 10107, [9]).
- "K" cold rolled electrical non-alloy and alloy steel sheet and strip delivered in the semi-processed state (DIN EN 10341, [10]).

Example: An electrical steel strip designated M330-35A is a non-oriented electrical steel strip in a finally annealed state, exhibits a remagnetization loss of 3.3 W/kg at 50 Hz and 1.5 T and has a thickness of 0.35 mm.

Materials for transformers and reactors (DIN IEC 740-2)
These materials comprise the alloy classes C21, C22, E11, E31 and E41 from the standard for soft-magnetic materials (DIN IEC 60404-8-6). The standard essentially contains the minimum values for core-sheet permeability for specified core-sheet sections (YEI, YED, YEE, YEL, YUI, and YM).

Materials for direct-current relays (DIN 17405)
Designation comprises a letter and number combination:
a) Code letter "R" (relay material).
b) Code letters for identifying alloying constituents:
Fe = unalloyed, Si = silicon steels,
Ni = nickel steels or alloys.
c) Code number for maximum coercive field strength.
d) Code letter for the stipulated delivery state: "U" = untreated, "GB" = malleable pre-annealed, "GT" = pre-annealed for deep-drawing, "GF" = final-annealed.

DIN IEC 60404-8-10 essentially contains the limit deviations for magnetic relay materials based on iron and steel. The designation code defined in this standard is as follows (example: M 80 TH):
– Code letter "M".
– Permitted maximum value for coercive field strength in A/m.
– Code letter for material composition: "F" = pure iron, "T" = steel alloy, "U" = unalloyed steel.
– Code letter for delivery state: "H" = hot-rolled, "C" = cold-rolled or cold-drawn.

Sintered metals for soft magnetic components (DIN IEC 60404-8-9)
Designation comprises a letter and number combination:
– Code letter "S" for sintered materials.
– Hyphen followed by the identifying alloy elements, i.e. Fe plus P, Si, Ni, or Co if necessary.
– The maximum permitted coercive field strength in A/m follows the second hyphen.

Soft magnetic ferrite cores (previously DIN 41280, [11])
Soft-magnetic ferrites are formed parts made of a sintered material with the general formula $MO{\cdot}Fe_2O_3$, where M is one or more of the bivalent metals Cd, Co, Ca, Mg, Mn, Ni, Zn.

Designation comprises a letter and number combination:
The various types of soft-magnetic ferrites are classified in groups according to nominal initial permeability and are designated by capital letters. Additional numbers may be used to further subdivide them into subgroups; these numbers have no bearing on material quality.

The coercive field strength H_c of soft ferrites is usually in the range of 4 to 500 A/m. Based on a field strength of 3,000 A/m, induction B is in the range of 350 to 470 mT.

Powder composite materials
Powder composite materials are not yet standardized, but are becoming increasingly more important. They consist of ferromagnetic metal powder (iron or an alloy) and an organic or inorganic grain-boundary phase as a "binder". They are manufactured in much the same way as sintered metals. The individual manufacturing stages are:
– Mixing the starting materials (metal powder and binder).
– Shaping by injection-molding, extruding or pressing.
– Heat treatment below the sinter temperature (< 600 °C).

Depending on the shaping process, binder type and the amount of binder used, it is possible to optimize the material to achieve high saturation polarization, higher permeability or high resistivity.

They are used primarily in fields in which all the above-mentioned characteristics are important, and where no excessively high demands are placed on mechanical strength and machinability. These fields currently consist of quick-acting actuators for diesel-injection engineering and high-speed small electric motors for motor vehicles.

Table 3: Materials for transformers and reactors
Core-sheet permeability for alloy classes C21, C22, E11, E31, and E41 for core-sheet section YEI1.

		Minimum core-sheet permeability μ_{lam} (min)											
IEC designation		C21-09 Thickness in mm				C22-13 Thickness in mm				E11-60 Thickness in mm			
		0.3 to 0.38	0.15 to 0.2			0.3 to 0.38				0.3 to 0.38	0.15 to 0.2	0.1	0.05
YEI 1	–10	630	630			1,000				14,000	18,000	20,000	20,000
	13	800	630			1,000				18,000	20,000	22,400	20,400
	14	800	630			1,000				18,000	22,400	22,400	22,400
	16	800	630			1,000				20,000	22,400	25,000	22,400
	18	800	630			1,000				22,400	25,000	25,000	22,400
	20	800	630			1,120				22,400	25,000	25,000	25,000
	22	800	630			1,120							
	25	800	630			1,120							
IEC designation		E11-100 Thickness in mm				E31-04 Thickness in mm				E31-06 Thickness in mm			
		0.3 to 0.38	0.15 to 0.2	0.1	0.05	0.3 to 0.38	0.15 to 0.2	0.1	0.05	0.3 to 0.38	0.15 to 0.2	0.1	0.05
YEI 1	–10	18,000	25,000	31,500	31,500	2,800	2,800	3,150	3,150	3,550	4,000	4,500	5,000
	13	20,000	28,000	35,500	35,500	2,800	3,150	3,150	3,550	4,000	4,500	5,000	5,000
	14	22,400	28,000	35,500	35,500	2,800	3,150	3,150	3,550	4,000	4,500	5,000	5,000
	16	25,000	31,500	35,500	35,500	2,800	3,150	3,150	3,550	4,500	4,500	5,000	5,000
	18	25,000	31,500	40,000	35,500	3,150	3,150	3,550	3,550	4,500	4,500	5,000	5,000
	20	28,000	35,500	40,000	40,000	3,150	3,150	3,550	3,550	4,500	5,000	5,000	5,000
IEC designation		E31-10 Thickness in mm				E41-02 Thickness in mm				E41-03 Thickness in mm			
		0.3 to 0.38	0.15 to 0.2	0.1	0.05	0.3 to 0.38	0.15 to 0.2	0.1	0.05	0.3 to 0.38	0.15 to 0.2	0.1	0.05
YEI 1	–10	5,600	6,300	5,600	6,300	1,600	1,800	1,800	2,000	2,000	2,240	2,500	2,240
	13	6,300	7,100	6,300	6,300	1,800	1,800	2,000	2,000	2,240	2,240	2,500	2,240
	14	6,300	7,100	6,300	7,100	1,800	1,800	2,000	2,000	2,240	2,240	2,500	2,240
	16	6,300	7,100	6,300	7,100	1,800	1,800	2,000	2,000	2,240	2,500	2,500	2,240
	18	7,100	7,100	6,300	7,100	1,800	1,800	2,000	2,000	2,240	2,500	2,500	2,240
	20	7,100	7,100	6,300	7,100	1,800	2,000	2,000	2,000	2,240	2,500	2,500	2,240

Table 4: Materials for DC relays

Material type		Alloying constituents	Density [1] ρ	Hardness [1]	Remanence [1]	Permeability [1]	Electr. resistivity [1]	Coercive field strength	Magnetic polarization in T (Tesla) min. at field strength H in A/m								Properties, application examples
Abbreviated Name	Material number	by mass %	g/cm³	HV	T (tesla)	μ_{max}	$\frac{\Omega \cdot mm^2}{m}$	A/m max.	20	50	100	200	300	500	1,000	4,000	
Unalloyed steels																	
RFe 160	1.1011	–	7.85	max. 150	–	–	0.15	160	–	–	–	–	1.15	1.30	–	1.60	Low coercive field strength. Direct-current relays and similar purposes.
RFe 80	1.1014				1.10	–	0.15	80	–	–	–	1.10	1.20	1.30	1.45	1.60	
RFe 60	1.1015				1.20	–	0.12	60	–	–	–	1.15	1.25	1.35	1.45	1.60	
RFe 20	1.1017				1.20	≈ 20,000	0.10	20	–	–	1.15	1.25	1.30	1.40	1.45	1.60	
RFe 12	1.1018				1.20		0.10	12	–	–	1.15	1.25	1.30	1.40	1.45	1.60	
Silicon steels																	
RSi 48	1.3840	2.5	7.55	130	0.50	–	0.42	48	–	–	0.60	–	1.10	1.20	–	1.50	
RSi 24	1.3843	–	–	–	1.00	≈ 20,000	–	24	–	–	1.20	–	1.30	1.35	–	1.50	
RSi 12	1.3845	4 Si	7.75	200	1.00	≈ 10,000	0.60	12	–	–	1.20	–	1.30	1.35	–	1.50	
Nickel steels and nickel alloys																	
RNi 24	1.3911	≈ 36 Ni	8.2	130 to 180	0.45	≈ 5,000	0.75	24	0.20	0.45	0.70	–	0.90	1.0	–	1.18	
RNi 12	1.3926	≈ 50 Ni	8.3	130 to 180	0.60	≈ 30,000	0.45	12	0.50	0.90	1.10	–	1.25	1.35	–	1.45	
RNi 8	1.3927	≈ 50 Ni	8.3	130 to 180	0.60	30,000 to 100,000	0.45	8	0.50	0.90	1.10	–	1.25	1.35	–	1.45	
RNi 5	2.4596	70 to 80 Ni	8.7	120 to 170	0.30	≈ 40,000	0.55	5	0.50	0.65	0.70	–	–	–	–	0.75	
RNi 2	2.4595	small quantities Cu, Cr, Mo	8.7	120 to 170	0.30	≈ 100,000	0.55	2	0.50	0.65	0.70	–	–	–	–	0.75	

[1] Standard values.

Table 5: Sintered metals for soft-magnetic components

Material Abbreviated name	Characteristic alloy elements (except Fe) % by mass	Sinter Density ρ_s	Porosity p_s	Maximum coercive field strength $H_{c(max)}$	Magnetic polarization in tesla (T) at field strength H in A/m				Maximum permeability	Vickers hardness	Volume Electrical resistivity ρ
	%	g/cm³	%	A/m	500	5,000	15,000	80,000	$\mu_{(max)}$	HV5	$\mu\Omega$m
S-Fe-175	–	6.6	16	175	0.70	1.10	1.40	1.55	2,000	50	0.15
S-Fe-170	–	7.0	11	170	0.90	1.25	1.45	1.65	2,600	60	0.13
S-Fe-165	–	7.2	9	165	1.10	1.40	1.55	1.75	3,000	70	0.12
S-FeP-150	≈ 0.45 P	7.0	10	150	1.05	1.30	1.50	1.65	3,400	95	0.20
S-FeP-130	≈ 0.45 P	7.2	8	130	1.20	1.45	1.60	1.75	4,000	105	0.19
S-FeSi-80	≈ 3 Si	7.3	4	80	1.35	1.55	1.70	1.85	8,000	170	0.45
S-FeSi-50	≈ 3 Si	7.5	2	50	1.40	1.65	1.70	1.95	9,500	180	0.45
S-FeNi-20	≈ 50 Ni	7.7	7	20	1.10	1.25	1.30	1.30	20,000	70	0.50
S-FeNi-15	≈ 50 Ni	8.0	4	15	1.30	1.50	1.55	1.55	30,000	85	0.45
S-FeCo-100	≈ 50 Co	7.8	3	100	1.50	2.00	2.10	2.15	2,000	190	0.10
S-FeCo-200	≈ 50 Co	7.8	3	200	1.55	2.05	2.15	2.20	3,900	240	0.35

Table 6: Soft magnetic ferrites

Ferrite type	Initial permeability [1] μ_i ±25%	Specific loss factor tan δ/μ_i [2] 10^{-6}	MHz	Specific Power loss [3] mW/g	Amplitude permeability [4] μ_a	Curie temperature [5][6] Θ_c °C	Frequency for $0.8 \cdot \mu_i$ [6] MHz	Properties, application examples
Materials in largely open magnetic circuits								
C 1/12	12	350	100	–	–	>500	400	Initial permeability. Compared to metallic magnetic materials high specific resistance (100 to $10^5\ \Omega \cdot m$, metals 10^{-7} to $10^{-6}\ \Omega \cdot m$), therefore low eddy-current losses. Telecommunications (coils, transformers).
D 1/50	50	120	10	–	–	>400	90	
F 1/250	250	100	3	–	–	>250	22	
G 2/600	600	40	1	–	–	>170	6	
H 1/1,200	1,200	20	0.3	–	–	>150	2	
Materials in largely closed magnetic circuits								
E 2	60 to 160	80	10	–	–	>400	50	
G 3	400 to 1,200	25	1	–	–	>180	6	
J 4	1,600 to 2,500	5	0.1	–	–	>150	1.5	
M 1	3,000 to 5,000	5	0.03	–	–	>125	0.4	
P1	5,000 to 7,000	3	0.01	–	–	>125	0.3	
Materials for power applications								
W 1	1,000 to 3,000	–	–	45	1,200	>180	–	
W 2	1,000 to 3,000	–	–	25	1,500	>180	–	

[1] Nominal values. [2] tan δ/μ_i denotes the frequency-dependent material losses at a low flux density ($B < 0.1$ mT).
[3] Losses at high flux density. Measured preferably at: $f = 25$ kHz, $B = 200$ mT, $\Theta = 100$ °C.
[4] Permeability when subjected to a strong sinusoidal magnetic field. Measured at: $f \leq 25$ kHz, $B = 320$ mT, $\Theta = 100$ °C.
[5] Curie temperature Θ_c in this table is the temperature at which initial permeability μ_i drops below 10% of its value at 25 °C.
[6] Standard values.

Permanent-magnet materials
(DIN 17410, replaced by DIN IEC 60404-8-1)
If chemical symbols are used in the abbreviated names of the materials, they refer to the primary alloying constituents of the materials. The numbers in the material abbreviation before the slash denote the $(BH)_{max}$ value in kJ/m^3 and those after the slash denote one tenth of the H_cJ value in kA/m (rounded values). Permanent magnets with binders are indicated by a final p.

Designation by material abbreviation or code number
Structure of code numbers
(DIN IEC 60404-8-1:2005-8):
Code letter (group)
+ 1st digit (material type),
+ 2nd digit of 0 (isotropic) or 1 (anisotropic),
+ 3rd digit (various quality levels).

R – Hard magnetic alloys,
such as R1: Aluminum-nickel-cobalt-iron-titanium alloys (AlNiCo).

S – Hard magnetic ceramic materials,
such as S1: Hard magnetic ferrite.

U – Combined hard magnetic materials,
such as
– U1: Combined aluminum-nickel-cobalt-iron-titanium magnets (AlNiCo).
– U2: Combined rare earth-cobalt magnets (RECo).
– U3: Combined neodymium-iron-boron magnets.
– U4: Combined hard ferrites.

Comparison of permanent and soft magnets
Figure 1 shows the magnetic characteristic value ranges of some crystalline materials in widespread use. The values of soft magnetic materials are compared to the values of hard magnets (permanent magnets).

Table 7: Permanent magnet materials
Bosch grades BTMT (not standardized)

Material	Density [1]	$(BH)_{max}$ [2]	Remanence [2]	Coercive field strength [2]	
	ρ		B_r	of flux density	of polarization
Abbreviated name				H_{CB}	H_{CJ}
	g/cm^3	kJ/m^3	mT	kA/m	kA/m
RBX HC 370		25	360	270	390
RBX HC 380		28	380	280	370
RBX 380 K		28	380	280	300
RBX 400		30	400	255	260
RBX 400 K	4.7 to 4.9	31	400	290	300
RBX HC 400		29	380	285	355
RBX 420		34	420	255	270
RBX 410 K		33	410	305	330
RBX HC 410		30	395	290	340
RBX 420 S		35	425	260	270
RBX HC 400 N		28	380	280	390

[1] Standard values. [2] Minimum values.

Table 8: Permanent magnet materials

Material			Chemical composition [1]							Density	$(BH)_{max}$ [2]	Remanence	Coercive field strength [2]		Rel. permanent permeability [1]	Curie temp. [1]	Temp. coeff. of polar. $TK(J_s)$ [1,3]	Temp. coeff. of coerciv. $TK(H_c)$ [1,3]	Manufacture, processing, applications
Abbreviated name	Materialnumber		% by weight										of flux density	of polarization					
	DIN	IEC	Al	Co	Cu	Nb	Ni	Ti	Fe	ρ [1]		B_r [2]	H_{CB}	H_{CJ}	μ_p	T_c			
										g/cm³	kJ/m³	mT	kA/m	kA/m		K	%K	%K	
Metallic magnets																			
Isotropic																			
AlNiCo 9/5	1.3728	R 1-0-3	11 to 13	0 to 5	2 to 4	–	21 to 28	0 to 1		6.8	9.0	550	44	47	4.0 to 5.0	1,030		+ 0.03	Manufacture: casting or sintering. For magnets with binders: pressing or injection molding. Processing: grinding. Application: max. 400 to 500 °C.
AlNiCo 18/9	1.3756	–	6 to 8	24 to 34	3 to 6	–	13 to 19	5 to 9	Rest	7.2	18.0	600	80	86	3.0 to 4.0	to	– 0.02		
AlNiCo 7/8p	1.3715	R 1-2-3	6 to 8	24 to 34	3 to 6	–	13 to 19	5 to 9		5.5	7.0	340	72	84	2.0 to 3.0	1,180		to – 0.07	
Anisotropic																			
AlNiCo 35/5	1.3761	–	8 to 9	23 to 26	3 to 4	0 to 1	13 to 16	–		7.2	35.0	1,120	47	48	3.0 to 4.5				
AlNiCo 44/5	1.3757	R 1-1-2	8 to 9	23 to 26	3 to 4	0 to 1	13 to 16	–		7.2	44.0	1,200	52	53	2.5 to 4.0	1,030			
AlNiCo 52/6	1.3759	–	8 to 9	23 to 26	3 to 4	0 to 1	13 to 16	–	Rest	7.2	52.0	1,250	55	56	1.5 to 3.0	to	– 0.02	+ 0.03 to – 0.07	
AlNiCo 60/11	1.3763	R 1-1-6	6 to 8	35 to 39	2 to 4	0 to 1	13 to 15	4 to 6		7.2	60.0	900	110	112	1.5 to 2.5	1,180			
AlNiCo 30/14	1.3765	–	6 to 8	38 to 42	2 to 4	0 to 1	13 to 15	7 to 9		7.2	30.0	680	136	144	1.5 to 2.5				
			Pt	Co															
PtCo 60/40	2.5210	R2-0-1	77 to 78	20 to 23						15.5	60	600	350	400	1.1	800	– 0.01 to – 0.02	– 0.35	
			V	Co	Cr	Fe													
FeCoVCr 11/2	2.4570	R 3-1-3	8 to 15	51 to 54	0 to 4					–	11.0	800	24	24	2.0 to 8.0	1,000	– 0.01	≈ 0	
FeCoVCr 4/1	2.4571	–	3 to 15	51 to 54	0 to 6	Rest				–	4.0	1,000	5	5	9.0 to 25.0				
RECo – Magnets of type $RECo_5$																			
RECo 80/80	–	R 5-1-1	Typically $MMCo_5$ (MM = Cer composition metal)							8.1	80	650	500	800	1.05	1,000	– 0.05	– 0.3	
RECo 120/96	–	R 5-1-2	Typically $SmCo_5$, typically (SmPr) Co_5							8.1	120	770	590	960	1.05	1,000	– 0.05	– 0.3	
RECo 160/80	–	R 5-1-3								8.1	160	900	640	800	1.05	1,000	– 0.05	– 0.3	
RECo – Magnets of type RE_2Co_{17}																			
RECo 165/50	–	R 5-1-11	Typically $MMCo_5$ (MM = Cer composition metal)							8.2	165	950	440	500	1.1	1,100	0.03	– 0.02	
RECo 180/90	–	R 5-1-13	Typically $SmCo_5$, typically (SmPr) Co_5							8.2	180	1,000	680	900	1.1	1,100	0.03	– 0.02	
RECo 190/70	–	R 5-1-14								8.2	190	1,050	560	700	1.1	1,100	0.03	– 0.02	
RECo 48/60p	–	R 5-1-3								5.2	48	500	360	600	1.05	1,000	– 0.05	– 0.3	

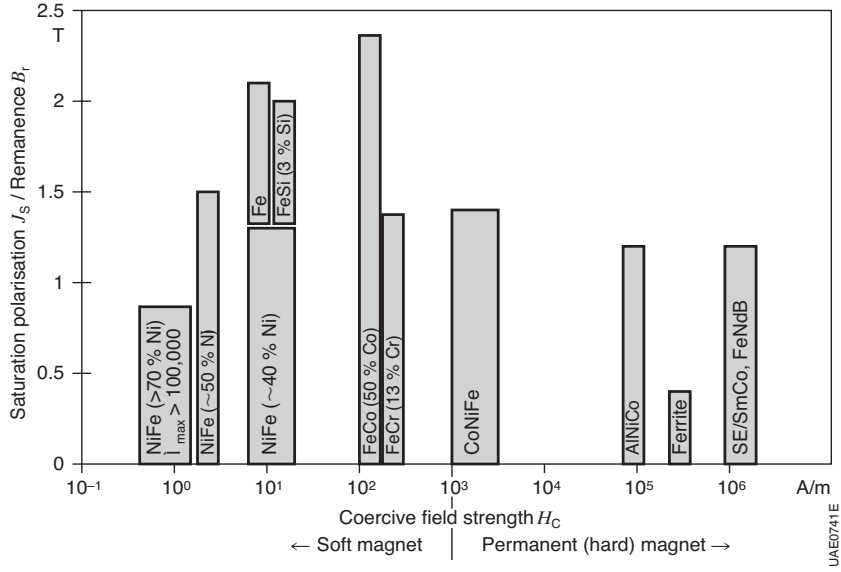
Material			Chemical composition[1]								Density	$(BH)_{max}$[2]	Remanence	Coercive field strength[2]		Rel. permanent permeability[1]	Curie temp.[1]	Temp. coeff. of polar.	Temp. coeff. of coerciv.	Manufacture, processing, applications
Abbreviated name	Material number		% by weight											of flux density	of polarization					
	DIN	IEC	Al	Co	Cu	Nb	Ni	Ti	Fe		ρ[1]		B_r[2]	H_{CB}	H_{CJ}	μ_p	T_c	$TK(J_s)$[1)3)]	$TK(H_c)$[1)3)]	
											g/cm³	kJ/m³	mT	kA/m	kA/m		K	%K	%K	
CrFeCo 12/4	–	R 6-0-1	(no data)								7.6	12	800	40	42	5.5 to 6.5	1,125	– 0.03	– 0.04	
CrFeCo 28/5	–	R 6-1-1									7.6	28	1,000	45	46	3 to 4	1,125	– 0.03	– 0.04	
REFe 165/170	–	R 7-1-1	(no data)								7.4	165	940	700	1,700	1.07	583	– 0.1	– 0.8	
REFe 220/140	–	R 7-1-6									7.4	220	1,090	800	1,400	1.05	583	– 0.1	– 0.8	
REFe 240/110	–	R 7-1-7									7.4	240	1,140	850	1,100	1.05	583	– 0.1	– 0.8	
REFe 260/80	–	R 7-1-8									7.4	260	1,180	750	800	1.05	583	– 0.1	– 0.8	

Material			Density[1]	$(BH)_{max}$[2]	Remanence[2]	Coercive field strength[2]		Rel. permanent permeability[1]	Curie-temp.[1]	Temp. coeff. of polarization[1]	Temp. coeff. of coer.[1]	Manufacture, processing, applications
Abbreviated name	Material number					of flux density	of polarization					
	DIN	IEC	ρ		B_r	H_{CB}	H_{CJ}	μ_p	T_c	$TK(J_s)$	$TK(H_c)$	
			g/cm³	kJ/m³	mT	kA/m	kA/m		K	%K	%K	
Ceramic magnets												
Isotropic												
Hard ferrite 7/21	1.3641	S 1-0-1	4.9	6.5	190	125	210	1.2	723	– 0.2	0.2 to 0.5	Manufacture: sintering. Plastic-bound magnets produced by pressing, injection molding, rolling, extruding. Processing: grinding.
Hard ferrite 3/18p	1.3614	S 1-2-2	3.9	3.2	135	85	175	1.1				
Anisotropic												
Hard ferrite 20/19	1.3643	S 1-1-1	4.8	20.0	320	170	190	1.1				
Hard ferrite 20/28	1.3645	S 1-1-2	4.6	20.0	320	220	280	1.1				
Hard ferrite 24/23	1.3647	S 1-1-3	4.8	24.0	350	215	230	1.1	723	– 0.2	0.2 to 0.5	
Hard ferrite 25/22	1.3651	S 1-1-5	4.8	25.0	370	205	220	1.1				
Hard ferrite 26/26	–	S 1-1-8	4.7	26.0	370	230	260	1.1				
Hard ferrite 32/17	–	S 1-1-10	4.9	32.0	410	160	165	1.1				
Hard ferrite 24/35	–	S 1-1-14	4.8	24.0	360	260	350	1.1				
Hard ferrite 9/19p	1.3616	S 1-1-3	3.4	9.0	220	145	190	1.1				
Hard ferrite 10/22p	–	S 1-2-3	3.5	10.0	230	165	225	1.1				

[1)] Standard values. [2)] Minimum values. [3)] In the range of 273 to 373 K.

References for magnetic materials
[1] DIN IEC 60404-8: Magnetic materials – Part 8: Specifications for individual materials.
8-1: Requirements for individual materials – Hard-magnetic materials (permanent magnets).
8-6: Requirements for individual materials – Soft-magnetic metallic materials.
8-9: Standard requirement for soft magnetic sintered metals.
8-10: Magnetic materials (iron and steel) for relay applications.
[2] DIN 17405: Soft magnetic materials for DC relay; technical terms of delivery.
[3] DIN IEC 60740-2: Laminations for transformers and inductors for use in telecommunication and electronic equipment – Part 2: Specification for the minimum permeabilities of laminations made of soft magnetic metallic materials.
[4] R. Boll: Weichmagnetische Werkstoffe – Einführung in den Magnetismus, VAC Werkstoffe und ihre Anwendungen [Soft magnetic materials – Introduction to magnetism, VAC materials and their applications]. 4th Edition, Publicis Corporate Publishing, 1990.
[5] L. Michalowsky, J. Schneider (ed.): Magnettechnik – Grundlagen, Werkstoffe, Anwendungen [Magnet technology – Foundations, materials, applications]. 3rd Edition, Vulkan-Verlag GmbH, 2006.
[6] Information Sheet 401, "Elektroblech und -band" [Electrical steel sheets and strips], Steel Information Center, Düsseldorf, 2005 edition.
[7] DIN EN 10027-1: Designation systems for steels – Part 1: Steel names.
[8] DIN EN 10106: Cold rolled non-oriented electrical steel sheet and strip delivered in the fully processed state.
[9] DIN EN 10107: Non-oriented electrical steel sheet and strip delivered in the fully processed state.
[10] DIN EN 10341: Cold rolled electrical non-alloy and alloy steel sheet and strip delivered in the semi-processed state.
[11] DIN 41280: Cores of soft magnetic oxides; material properties

Figure 1: Comparison of permanent and soft magnets



Nonmetallic inorganic materials

These materials are characterized by ionic bonds (e.g. ceramic materials), mixed (heteropolar/homopolar) bonds (e.g. glass), or homopolar bonds (e.g. carbon). These kinds of bonds, in turn, are responsible for several characteristic properties: generally poor thermal conductivity and poor electric conductivity (the latter increases with temperature), poor luminous reflectance, brittleness, and thus almost complete unsuitability for cold forming.

Ceramics
Ceramics are at least 30 % crystalline in nature; most ceramics also contain amorphous components and pores. Their manufacture is similar to that of sintered metals, however nonmetallic powders or powder mixtures are used; sintering at temperatures generally higher than 1,000 °C gives ceramics their characteristic properties. Ceramic structural parts are sometimes also shaped at high temperatures or even by a melting process, with subsequent crystallization.

Table 1 gives an overview of ceramic materials and their properties.

Glass
Glass is viewed as under-cooled, frozen liquid. Its atoms are only in a short-range order. It is regarded as amorphous. Molten glass turns to solid glass at transformation temperature T_g (T_g is derived from the former designation glass formation temperature). The transformation temperature is dependent on a variety of parameters and therefore not clearly determined (better: transformation range).

Composite materials

Composite materials consist of at least two physically or chemically different components. These components must be tightly bonded together at a specific boundary layer. Under these conditions, it is possible to bond many materials together. The resulting composite material exhibits properties that neither of the original materials has. A distinction is made between the following:
– Particle composite materials: e.g. powder-filled resins, hard metals, plastic-bonded magnets, cermets.
– Laminated composite materials: Such as thermostatic bimetals, sandwich panels.
– Fiber composite materials: Such as glass fiber or carbon fiber-reinforced plastics.

References for Nonmetallic inorganic materials

[1] DIN EN 623: Advanced technical ceramics; monolithic ceramics; general and textural properties; Part 2: determination of density and porosity.
[2] DIN EN 843: Advanced technical ceramics – Mechanical properties of monolithic ceramics at room temperature. Part 1: Determination of flexural strength. Part 2: Determination of Young's modulus, shear modulus and Poisson's ratio.
[3] DIN EN 993: Methods of test for dense shaped refractory products – Part 5: Determination of cold crushing strength.
[4] DIN EN 821: Advanced technical ceramics – Monolithic ceramics – Thermophysical properties. Part 1: Determination of thermal expansion. Part 2: Determination of thermal diffusivity by the laser flash (or heat pulse) method. Part 3: Determination specific heat capacity.
[5] DIN EN 60672-1: Ceramic and glass insulating materials – Part 1: Terms and classification.
[6] www.keramikverband.de.
[7] www.matweb.com.

Table 1: Ceramic materials

Materials	Composition	ρ [1] g/cm³	σ_{bB} [2] MN/m²	σ_{dB} [3] MN/m²	E [4] GN/m²	α_t [5] 10^{-6}/K	λ [6] W/mK	c [7] kJ/kg · K	ρ_D [8] Ω · cm	ε_r [9]	tan ρ [10] 10^{-4}
Aluminum nitride	AlN > 97 %	3.3	250 to 350	1,100	320 to 350	5.1	100 to 220	0.8	> 10^{14}	8.5 to 9.0	3 to 10
Aluminum oxide	Al_2O_3 > 99 %	3.9 to 4.0	300 to 400	3,000 to 4,000	380 to 400	7.2 to 8.6	20 to 40	0.8 to 0.9	> 10^{11}	8 to 10	2
Aluminum titanate	$Al_2O_3 \cdot TiO_2$	3.0 to 3.7	25 to 40	450 to 550	10 to 30	0.5 to 1.5	< 2	0.7	> 10^{11}	–	–
Beryllium oxide	BeO > 99 %	2.9 to 3.0	250 to 320	1,500	300 to 340	8.5 to 9.0	240 to 280	1.0	> 10^{14}	6.5	3 to 5
Boron carbide	B_4C	2.5	300 to 500	2,800	450	5.0	30 to 60	–	10^{-1} to 10^{2}	–	–
Cordierite, e.g. C 410, 520 [11]	$2MgO \cdot 2Al_2O_3 \cdot 5SiO_2$	1.6 to 2.3	5 to 100	300	6 to 60	2.0 to 5.0	1.3 to 2.5	0.8	> 10^{11}	5.0	70
Graphite	C > 99.7 %	1.5 to 1.8	5 to 30	20 to 50	5 to 15	1.6 to 4.0	100 to 180	–	10^{-3}	–	–
Porcelain, e.g. C 110 – 120 (non-glazed)	Al_2O_3 30 to 35 % rest SiO_2 + glass phase	2.2 to 2.4	20 to 100	500 to 550	50	4.0 to 6.5	1.2 to 2.6	0.8	10^{11}	6	120
Silicon carbide pressureless-sintered SSiC	SiC > 98 %	3.1 to 3.2	400 to 600	> 1,200	400	4.0 to 4.5	90 to 120	0.8	10^{3}	–	–
Silicon carbide hot-pressed HPSiC	SiC > 99 %	3.1 to 3.2	450 to 800	> 1,500	420	4.0 to 4.5	100 to 120	0.8	10^{3}	–	–
Silicon carbide reaction-sintered SiSiC	SiC > 90 % + Si	3.0 to 3.1	300 to 400	> 2,200	380	4.2 to 4.8	100 to 160	0.8	10 to 100	–	–
Silicon nitride gas-pressure-sintered GPSN	Si_3N_4 > 90 %	3.2	800 to 1,400	> 2,500	300	3.2 to 3.5	30 to 45	0.7	10^{12}	–	–
Silicon nitride hot-pressed HPSN	Si_3N_4 > 95 %	3.2	600 to 900	> 3,000	310	3.2 to 3.5	30 to 45	0.7	10^{12}	–	–

Materials	Composition	ρ [1] g/cm³	σ_{bB} [2] MN/m²	σ_{dB} [3] MN/m²	E [4] GN/m²	α_t [5] 10^{-6}/K	λ [6] W/mK	c [7] kJ/kg · K	ρ_D [8] Ω · cm	ε_r [9]	tan ρ [10] 10^{-4}
Silicon nitride reaction-sintered RBSN	Si_3N_4 > 99 %	2.4 to 2.6	200 to 300	< 2,000	140 to 160	2.9 to 3.0	15 to 20	0.7	10^{14}	–	–
Steatite, e.g. C 220, 221	SiO_2 55 to 65 % MgO 25 to 35 % Al_2O_3 2 to 6 % Alkali oxide < 1.5 %	2.6 to 2.9	120 to 140	850 to 1,000	80 to 100	7.0 to 9.0	2.3 to 2.8	0.7 to 0.9	> 10^{11}	6	10 to 20
Titanium carbide	TiC	4.9	–	–	320	7.4	30	–	$7 \cdot 10^{-5}$	–	–
Titanium nitride	TiN	5.4	–	–	260	9.4	40	–	$3 \cdot 10^{-5}$	–	–
Titanium dioxide	TiO_2	3.5 to 3.9	90 to 120	300	–	6.0 to 8.0	3 to 4	0.7 to 0.9	–	40 to 100	8
Zirconium dioxide partially stabilized, PSZ	ZrO_2 > 90 % rest Y_2O_3	5.7 to 6.0	500 to 1.000	1,800 to 2,100	140 to 210	9.0 to 11.0	2 to 3	0.4	10^8	–	–
Standards for test procedures		DIN EN 623 Part 2 [1]	DIN EN 843 Part 1 [2]	DIN EN 993 Part 5 [3]	DIN EN 843 Part 2 [2]	DIN EN 821 Part 1 [4]	DIN EN 821 Part 2 [4]	DIN EN 821 Part 3 [4]			

The characteristic values for each material can vary widely, depending on the raw material, composition, and manufacturing process. The material data relate to the information provided by various manufacturers.
The designation "KER" corresponds to DIN EN 60-672-1 [5].
Further detailed information on ceramic materials, property tables, applications, etc., can be found on the relevant internet pages, e.g. Informationszentrum Technische Keramik (center of information for technical ceramics) [6], or MatWeb Material Property Data [7]

[1] Density. [2] Flexural strength. [3] Cold compressive strength. [4] Modulus of elasticity. [5] Coefficient of thermal expansion RT to 1,000 °C.
[6] Thermal conductivity at 20 °C. [7] Specific heat. [8] Electrical resistance at 20 °C and 50 Hz. [9] Relative permittivity.
[10] Dielectric loss factor at 25 °C and 10 MHz. [11] Properties greatly dependent on porosity set specifically for the application.

Plastics

Plastics are still a relatively "young" material group, the importance of which has continuously grown since the middle of the twentieth century. Products made of plastics play a large role in the various areas of our society today. Thus plastics are used as material in areas such as supply, medical and electrical technology, for packaging, household appliances and consumer goods.

Specifically in the automotive industry, plastics have been able to gain increased use more than any other material group (Figure 1). The automotive industry is often the driver behind technological innovations for plastic.

However, the "saturation limit" for using plastics is far from having been reached. We continue to find applications in which conventional materials such as metal are being replaced by plastics.

The increasing use of plastics is mostly due to their less expensive processing options compared to other materials and to the beneficial properties of this "made-to-measure material group". In addition, new material and process developments are opening new markets and simultaneously offering enormous potential for new, innovative products made of plastic. A strong increase in the plastic content is forecast particularly due to increasing electromobilization. The greatest driver is the savings in weight compared to metallic materials.

Plastics is the general umbrella term for polymer materials. A significant characteristic is their macromolecular structure.

Figure 1: Average percentage of individual material groups in a mid-class car [1]

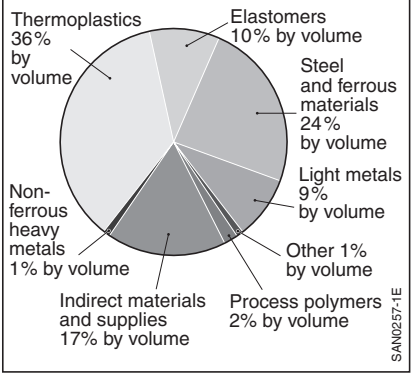
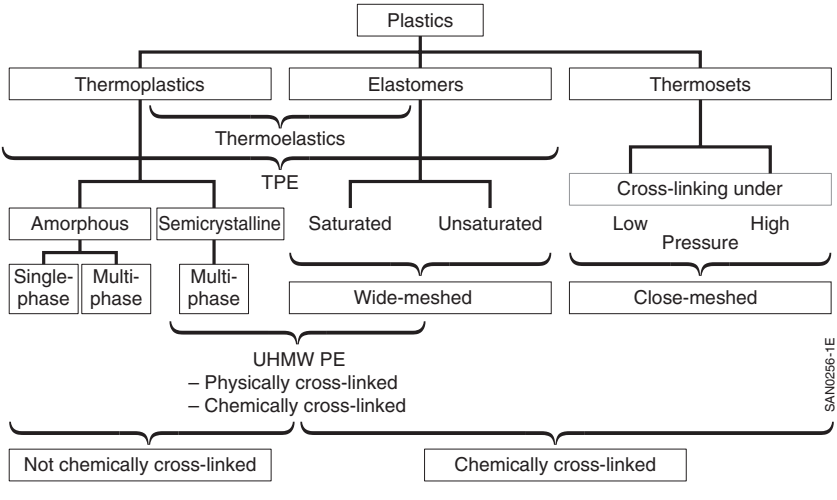


Figure 2: Classification of plastics based on [2]



Plastics are classified into thermoplastics, thermosets, elastomers and thermoplastic elastomers (Figure 2), which are discussed in more detail in the following.

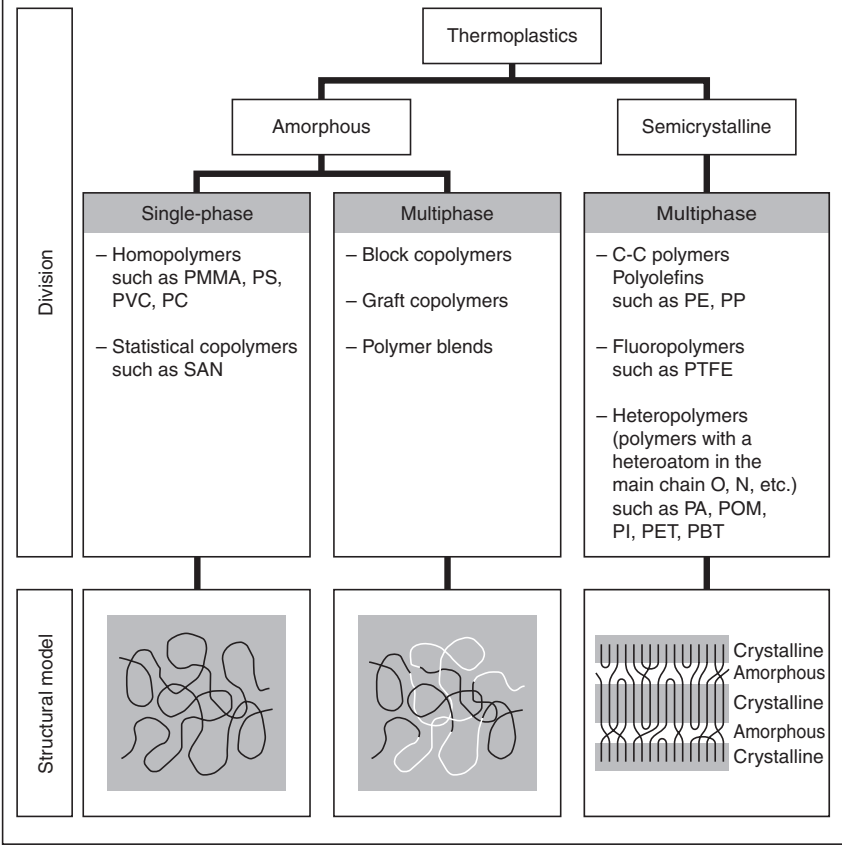
Thermoplastics

An outstanding characteristic of thermoplastics is the non-cross-linked structure between the macromolecules. This enables repeated plasticity or processability above their service temperature in the melting range. Only thermoplastics can be welded.

The material class of the thermoplastics can be further classified into amorphous thermoplastics and semicrystalline thermoplastics [3]. Semicrystalline thermoplastics are composed of amorphous and semicrystalline macromolecular structures. Unlike amorphous thermoplastics, therefore, they exist in multiple phases. Even amorphous thermoplastics can exist in multiple phases if they are modified during their synthesis into the form of a copolymer. The correlations are described in Figure 3 (based on [2]). The block copolymers existing in multiple phases form the link to the material class of thermoplastic elastomers.

The number of manufacturers and commercial types in the area of thermoplastics is very large. Most polymers come with or without different kinds and percentages of fillers and reinforcing materials. Therefore

Figure 3: Classification of thermoplastics based on [2]



the range of properties of thermoplastics available on the market is very large. Table 1 provides an overview of names, codes and example applications of widespread thermoplastics.

Mechanical properties of thermoplastics
Compared to other construction materials, thermoplastics have exceptional viscoelastic and viscoplastic deformation behavior. Consequently, the mechanical behavior is strongly influenced by temperature, load exposure period and loading rate (Figure 4). Amorphous thermoplas-

Table 1: Chemical name and properties of thermoplastics

Code	Chemical name	Description of properties; application examples
ABS	Acrylnitrile-butadien-styrene	High gloss, some transparent types; impact-resistant housing parts
PA 11, 12	Polyamide 11, 12	Tough, hard, and resistant to abrasion, low coefficient of friction, good sound absorption, approx. 1 to 3% water absorption required for good toughness; PA 11, 12 have much lower water absorption
PA6	Polyamide 6	
PA66	Polyamide 66	
PA6-GF	Polyamide 6 + GF	Impact-resistant machine housings
PA66-GF	Polyamide 66 + GF	
PA6T/6I/66-GF	Polyamide 6T/6I/66 +GF	Rigid machine housings and components even at high temperatures, low water absorption as standard PA
PA6/6T-GF	Polyamide 6/6T + GF	
PAI	Polyamid-imide	Components that are subjected to mechanical or electrical stress, favorable wear characteristics
PBT	Polybutylene terephthalate	Wear-resistant, chemically resistant, hydrolysis in water over 70 °C, very good electrical properties
PBT-GF	Polybutene terephthalate + GF	
PC	Polycarbonate	Tough and rigid over a broad temperature range, components have high rigidity
PC-GF	Polycarbonate + GF	
PE	Polyethylene	Acid-resistant containers and pipes, films
PET	Polyethylene terephthalate	Wear-resistant, chemically resistant, hydrolysis in water over 70 °C
PET-GF	Polyethylene terephthalate + GF	
LCP-GF	Liquid crystal polymers + GF	High dimensional stability under heat, low weld line strength, extremely thin-walled components, very anisotropic
PESU-GF	Polysulfone + GF	High continuous service temperature, low dependency of properties on temperature, not resistant to fuel and alcohol at higher temperatures, dimensionally stable components

Table 1 (continued): Chemical name and properties of thermoplastic plastics

Code	Chemical name	Description of properties; application examples
PEEK-GF	Polyetheretherketone + GF	High-strength components for high temperatures, good antifriction properties, and low wear; very expensive
PMMA	Polymethylmethacrylate	Transparent and in many colors, weatherproof
POM	Polyoxymethylene	Sensitive to acid-induced stress cracking; precision moldings
POM-GF	Polyoxymethylene + GF	
PPE+SB	Polyphenylene ether + SB	Resistant to hot water, flame retardant
PPS-GF	Polyphenylene sulfide + GF	High resistance to heat and media, inherently flame-retarding; underhood components
PP	Polypropylene	Household goods, battery cases, cover hoods, fan impellers (reinforced types)
PP-GF	Polypropylene + GF	
PS	Polystyrene	Transparent and in many colors
PSU-GF	Polysulfone + GF	Low dependence of properties on temperature, not resistant to fuel and alcohol
PVC-P	Polyvinyl chloride (plasticized)	Artificial leather, flexible caps, cable insulation, tubes/hoses, seals
PVC-U	Polyvinyl chloride (unplasticized)	Weatherproof exterior parts, pipes
SAN	Styrene-acrylonitrile	Molded parts with good chemical resistance, also transparent
SB	Styrene-butadiene	Impact-resistant housing parts for many applications
SPS-GF	Syndiotactic polystyrene	Low-warpage, brittle, high mold temperatures required for processing
PI	Polyimide	High resistance to heat and radiation, can be processed only by compacting and sintering
PTFE	Polytetrafluorethylene	Strong dependence of rigidity on temperature, high resistance to heat, aging and chemicals, can be processed only by compacting and sintering

tics and semicrystalline thermoplastics behave differently. This is due to the different path of the shear modulus curve and the different transition areas associated with this. They are called glass transition temperature T_g and melting range (Figure 5, based on [2] and [3]). With thermoplastics, the combination of load and higher temperature leads to creep or relaxation due to the viscoelastic and viscoplastic material behavior. In Figure 6, the creep and relaxation behavior of thermoplastics is compared to the material behavior of metals [2].

Figure 4: Depiction of the effect of different test speeds and temperatures on the stress-strain behavior of thermoplastics
a) Dependency on the test speed,
b) Dependency on the temperature.

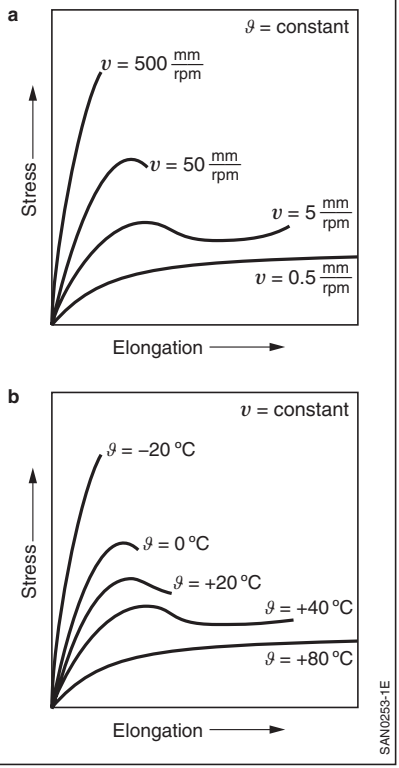


Table 2 and Table 3 show the mechanical properties of amorphous and semicrystalline materials. The values reproduce reference values from literature ([2] and [4]). The values may differ depending on the manufacturer and composition of filler or reinforcing material.

Chemical properties of thermoplastics
The chemical behavior of thermoplastics is determined by the structure of the macromolecules from which they are built. Polar plastics are attacked by polar solvents; non-polar plastics are attacked by non-polar solvents. Substances with a low molecular weight can migrate through solid thermoplastics (permeation).

Permeation of materials with a low molecular weight can trigger formation of stress cracks, much like with metals (referred to here as stress corrosion cracking). Suitable thermoplastics can be selected for a large number of application areas by selecting the right type of thermoplastic, appropriate shaping for the material and article, and optimum production parameters. Table 4 shows the chemical properties of some thermoplastics. The values are reference values from literature ([2], [4], [6]). Since the chemical behavior is difficult to estimate in some cases, consulting the plastics manufacturer or conducting your own measurements is advised.

Durability and aging of thermoplastics
Aging includes all irreversible changes to a material over time [7]. Aging processes change the properties of thermoplastics during a certain period of time. The causes of aging are distinguished between inner (such as residual stress, limited miscibility of additives) and external (energy input from heat and radiation, temperature change, mechanical loading, chemical influences). These causes lead to aging processes that manifest themselves in signs of aging and have a visible or measurable effect on thermoplastics. Examples of this are swelling, aftershrinkage, discoloration and alteration of the mechanical properties, such as brittleness.

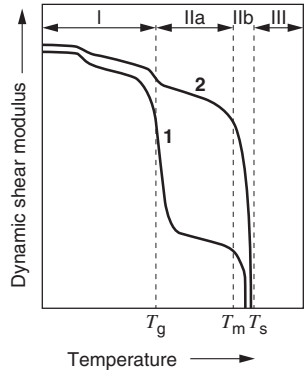
Figure 5: Temperature dependence of the dynamic shear modulus
1 Amorphous thermoplastics,
2 Semicrystalline thermoplastics.
T_g Glass transition temperature,
T_m Melt or flow temperature,
T_s Crystalline melting temperature.

Amorphous thermoplastics

Area I:
Glassy state,
Energy-elastic behavior;
Scope

Area IIa:
Entropy-elastic behavior
(viscoplastic),
Hot working range

Area III:
Viscous flow behavior,
Primary forming
and welding area



Semicrystalline thermoplastics

Area I:
Glassy state,
Energy-elastic behavior,
Amorphous areas frozen

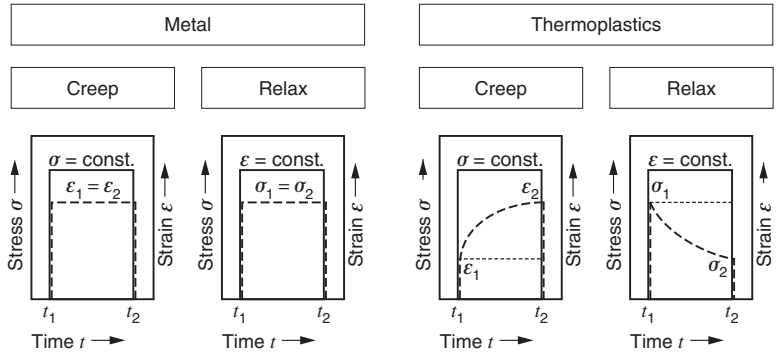
Area IIa:
Amorphous parts
Thermoelastic,
Semicrystalline
parts rigid;
Scope

Area IIb:
Crystallites begin
to fuse,
Hot working range

Area III:
Viscous flow behavior,
Primary forming
and welding area

SAN0254-E

Figure 6: Comparison of the influence of time on the mechanical behavior of metals and thermoplastics
T_R Recrystallization temperature, T_g Glass transition temperature.



Influence of t is relatively low as long as $T < T_R$

Influence of t is to be noted particularly if $T > T_g$
The smaller the E-modulus, the greater the tendency to creep

SAN0255-1E

Table 2: Mechanical properties of amorphous thermoplastics

Code	E module [N/mm²]	Tensile strain at yield or tensile stress at break [N/mm²]	Tensile strain at yield or Elongation at fracture [%]	Glass transition temperature [°C]	Continuous service temperature [°C]
ABS	1,300 to 2,700	32 to 45 (y)	15 to 30 (b)	80 to 110	75 to 85
PAI-GF30	10,800	205 (b)	7 (b)	240 to 275	260
PC	2,100 to 2,400	56 to 67 (y)	100 to 130 (b)	150	130
PEI-GF30	9,000	160 (y)	3 (b)	215	170
PESU-GF20	5,700 to 7,500	105 to 130 (b)	2.5 to 3.2 (b)	220 to 225	160 to 200
PMMA	1,600 to 3,600	50 to 77	2 to 10	110	65 to 90
PS	3,200 to 3,250	45 to 65	3 to 4	95 to 100	60 to 80
PSU-GF20	6,200 to 7,000	100 to 115 (b)	2 to 3 (b)	180 to 190	160 to 180
PVC-E	2,000 to 3,000	50 to 60 (y)	10 to 50 (b)	80	65
PVC-S	2,000 to 3,000	50 to 60 (y)	10 to 20 (b)	85	65
SAN	3,600	75 (b)	5 (b)	110	85
PI	3,000* to 3,200*	75 to 100 (b)	n/s	250 to 270	260

* Flexural-E-modulus,
y Value at first maximum value during the tensile test,
b Value at which the specimen breaks,
n/s Not specified.

More physical properties of the thermoplastics

Thermoplastics have good insulation properties against electricity and heat. Compared to other materials, thermoplastics have significantly larger and anisotropic coefficients of thermal expansion ([2], [4], [5]). Some thermoplastics, above all polyamides, absorb a significant amount of water, which results in changes to the mechanical and physical properties as well as the component dimensions.

Processing of thermoplastics

Thermoplastics are usually made available by raw material producers in granular form as sacked goods and can be processed in commercially available injection molding machines and extruders. Since thermoplastics are deformable, they can also be processed using deep-draw or compression methods. Since thermoplastics are meltable and solidify again, they can be also be welded during the production process and can be recycled to a certain extent.

Raw material producers usually provide recommendations for processing thermoplastics.

Use of thermoplastics

Thermoplastics are used in a wide variety of industries and in very diverse applications. Thermoplastics are used in the packaging industry, construction industry, consumer goods area (household appliances, toys, sports equipment), medical technology and aerospace technology. In the automotive industry, thermoplastics are being used very successfully in the interior, exterior and powertrain areas.

Table 3: Mechanical properties of semicrystalline thermoplastics

Code	E module [N/mm²]	Tensile strain at yield or tensile stress at break [N/mm²]	Tensile strain at yield or Elongation at fracture [%]	Glass transition temperature [°C]	Continuous service temperature [°C]
PA 11	1,370 (k)	42 (k, y)	5 (k, y)	49 (tr)	70 to 80
PA 11-GF30	7,300 (k)	134 (k, b)	6 (k, b)	49 (tr)	70 to 80
PA 12	1,600 (tr) / 1,100 (k)	50 (tr, y) / 40 (k, y)	5 (tr, y) / 12 (k, y)	49 (tr)	70 to 80
PA 12-GF30	8,000 (tr) / 7,500 (k)	130 (tr, b) / 120 (k, b)	6 (tr, b) / 6 (k, b)	49 (tr)	70 to 80
PA 6	3,000 (tr) / 1,000 (k)	85 (tr, y) / 40 (k, y)	4.5 (tr, y) / 20 (k, y)	60 (tr)	80 to 100
PA 6-GF30	9,500 (tr) / 6,200 (k)	185 (tr, y) / 115 (k, y)	3.5 (tr, y) / 8 (k, y)	60 (tr)	100 to 130
PA 66	3,000 (tr) / 1,100 (k)	85 (tr, y) / 50 (k, y)	4.4 (tr, y) / 20 (k, y)	70 (tr)	80 to 120
PA 66-GF30	10,000 (tr) / 7,200 (k)	190 (tr, y) / 130 (k, y)	3 (tr, y) / 5 (k, y)	70 (tr)	100 to 130
PAEK-GF30	10,600	168 (b)	2.3 (b)	158	240 to 250
PBT	2,500	60 (y)	3.7 (y) / >50 (b)	60	100
PBT-GF30	10,000	135	2.5 (b)	60	150
PE-LD	200 to 500	8 to 23	300 to 1,000	–30	60 to 75
PET	2,800	80 (y)	4 (y) / 12 (b)	98	100
PET-GF35	14,000	150 (b)	1.5 (b)	98	100
PEEK-GF30	9,700	156 (b)	2 (b)	145	240
POM (H)	3,200	67 to 72 (y)	25–70 (b)	–60	90 to 110
POM (CoP)	2,800	65 to 70 (y)	25–70 (b)	–60	90 to 110
PPS-GF	14,700	195 (b)	1.9 (b)	85 to 95	200 to 240
PP	1,100 to 1,300	30 (y)	20–800 (b)	–10 to 0	100
PTFE	408	25 to 36 (y)	350–550 (b)	127	260

k Conditioned,
tr Dry,
y Value at first maximum value during the tensile test,
b Value at which the specimen breaks.

Table 4: Chemical properties of thermoplastics

Code	Gasoline	Benzene	Diesel	Alcohol	Mineral oil	Brake fluid	Water cold/hot
ABS	+	–	+	+	+	–	+/+
PA 11	+	+	+	+	+	+	+/O
PA 12	+	+	+	+	+	n/s	+/O
PA6	+	+	+	+	+	+/O	+/O
PA6-GF30	+	+	+	+	+	+/O	+/O
PA66	+	+	+	+	+	+/O	+/O
PA66-GF30	+	+	+	+	+	+/O	+/O
PAI-GF	+	n/s	n/s	+	+	n/s	+/–
PBT	+	O	+	+	+	+	+/–
PBT-GF30	+	O	+	+	+	+	+/–
PC	+	–	O	O	+	n/s	+/O
PC-GF	+	–	O	O	+	n/s	+/O
PE	O	O	+	+	+	n/s	+/+
PET	+	+	+	+	+	n/s	+/–
PET-GF30	+	+	+	+	+	n/s	+/–
PESU-GF	+	–	+	O	+	–	+/O
PEEK-GF	+	n/s	n/s	+	n/s	n/s	+
PI	+	O	+	+	+	n/s	+/O
PMMA	+	O	+	–	+	n/s	+/+
POM (H)	+	+	+	+	+	n/s	+/+
POM-GF (H)	+	+	+	+	+	n/s	+/+
POM (CoP)	+	+	+	+	+	+	+
POM-GF (CoP)	+	+	+	+	+	+	+
PPS-GF40	+	+	+	+	+	+	+/+
PP	O	O	+	+	+	n/s	+/+
PP-GF30	O	O	+	+	+	n/s	+/+
PS	–	–	O	+	O	–	+/+
PSU	+	–	+	O	+	–	+/O
PTFE	+	+	+	+	+	n/s	+/+
PVC-P	–	–	O	–	O	n/s	+/O
PVC-U	+	–	+	+	+	n/s	+/O
SAN	–	–	O	+	+	–	+/+

+ Good resistance O Limited resistance – No resistance n/s Not specified

Thermoplastic elastomers
Thermoplastic elastomers (TPE) form their own plastic material class between thermoplastics and elastomers. They can be processed in a purely physical process that combines high shear forces, heat effects and subsequent cooling (for example, during injection molding or extrusion). Although no chemical cross-linking is required by a time-consuming and high-temperature vulcanization process such as with elastomers, the manufactured parts do in fact have rubber-elastic properties due to their special molecular structure. Renewed influence of heat and shear force causes the material to melt and deform. At the same time, however, this means that the thermoplastic elastomers can withstand far less thermal and dynamic loads than standard elastomers. The thermoplastic elastomers are not a "successor product" of conventional elastomers, but rather an interesting supplement, which combines the processing benefits of thermoplastics with the material properties of the elastomers ([2], [5], [8], [9], [10]).

Types of thermoplastic elastomers
"Thermoplastic elastomers" is the umbrella term for a whole series of different materials. They are always created through blends or block copolymers.

The blends are alloys made of a plastic matrix and a soft elastomer material. Block copolymers are molecular chains with different segments that congregate into hard and soft areas during cooling. Based on DIN EN ISO 18064 [11], the following subdivision can be made.

– TPO: Thermoplastic elastomers on an olefin basis, primarily PP/EPDM, such as Santoprene (Exxon Mobil).
– TPV: Cross-linked thermoplastic elastomers on an olefin basis, primarily PP/EPDM, such as Sarlink (DSM) and Forprene (SoFter).
– TPU: Thermoplastic elastomers on a urethane basis, such as Desmopan (Bayer) and Elastollan (BASF).
– TPC: Thermoplastic copolyester elastomers, such as Hytrel (DuPont) and Riteflex (Ticona).
– TPS: Styrene block copolymers (SBS, SEBS, SEPS, SEEPS and MBS), such as thermoplastic (Kraiburg TPE).
– TPA: Thermoplastic copolyamides, such as PEBAX (Arkema).

Properties of thermoplastic elastomers
The number of diverse commercial types with different properties on the market is very large. They, like thermoplastics, are usually made available by raw material producers in granular form as sacked goods.

Thermoplastic elastomers can be processed very well using injection molding and extrusion processes, since they pass through the plastic, molten state. They can be manufactured in all hardnesses from 5 Shore A to over 70 Shore D. The hardness and the compression set (DVR) are essential characteristics when using thermoplastic elastomers as sealants. Their thermal stability in particular is usually lower than that of elastomers. The maximum continuous service temperatures that can currently be attained for thermoplastic elastomers are approximately 150 °C.

Adhesion to nearly all technical thermoplastics can be achieved through modification. Their flowability as well as density, optics, scratch resistance and other properties can also be adjusted through compounding with a wide variety of fillers and additives.

Application of thermoplastic elastomers
Thermoplastic elastomers find a wide variety of applications in various industries while fulfilling the industry-standard requirements. Thus they are used in the automotive industry for control elements in the interior as well as for window trim in the exterior and for seals close to the engine. Furthermore, they are used in the industrial sector, for example, for tool handles or cable jackets. In the consumer sector, thermoplastic elastomers are found in toys, sports equipment, packaging, and personal care supplies such as toothbrushes and shavers. There are special compounds that satisfy the high requirements even for medical applications. They are used, for example, for drip chambers, seals and medical hoses.

Elastomers
Elastomers (or rubber compounds) are dimensionally stable, but elastically deformable plastics. The essential characteristic of these materials is that they can be stretched to at least twice their length. However, when the tensile load or compressive load is removed, they return to their initial state. This unique ability to reset is also referred to as rubber elasticity.

Elastomers are created from loose, wide-meshed and three-dimensional cross-linking of amorphous preliminary products (rubber). This loose fixation of polymer chains through chemical bonds leads to typical elastic behavior above the glass transition temperature T_g. This value usually lies significantly below 0 °C, and therefore below the service temperature of rubber compounds.

On the other hand, if the cross-linking is made solid and close-meshed, we speak of a thermoset. Cross-linked elastomers (just like thermosets) cannot be melted; that is, they decompose at high temperatures without melting.

The starting point for manufacturing elastomer materials is either natural or synthetic rubber. The rubber is mixed with various additives such as fillers, plasticizers, cross-linking chemicals, anti-aging agents and processing aids; then they are cross-linked under the influence of temperature.

During the cross-linking (vulcanization) a chemical reaction is underway, which usually takes place during the molding process under the influence of temperature (150 to 210 °C) and pressure. Not until after the vulcanization does the material have its rubber-elastic and mechanical properties, such as the desired hardness, tensile strength and elongation at break ([9], [12], [13], [14], [15]).

Classification of elastomers
Elastomers is an umbrella term for a whole material class whose individual materials can have widely differing properties. They can be classified into groups; based on the DIN ISO 1629 standard [16], the following subdivision can be made.
- R class: Rubbers with an unsaturated carbon chain, such as NR, NBR and SBR.
- M class: Rubbers with a saturated carbon chain, such as ACM, EPDM and FKM.
- O class: Rubbers with carbon and oxygen in the polymer chain, such as ECO.
- U class: Rubbers with carbon, oxygen and nitrogen in the polymer chain, such as polyurethane elastomers AU and EU.
- Q class: Rubbers with silicon and oxygen in the polymer chain, such as silicone elastomers VMQ.

Here, note that the ISO name (for example, NR for Natural Rubber) merely refers to the base rubber.

Properties of elastomers
General basic properties such as thermal and media resistance are already largely determined by the base rubber being used. These can be changed only within certain (very limited) boundaries. Yet within an elastomer class, such as NBR, there is a wide variety of different mixtures that can have significantly different specific properties such as hardness, strength behavior and elasticity. In addition, component suppliers often use their own recipes for their materials and the rubber compounds are often specially designed for the requirements of an application. Unlike thermoplastics or thermoplastic elastomers, non-cross-linked elas-

tomer mixtures are available on the free market only in a few cases and are not offered also in granular form as sacked goods or the like.

Tables 5 and 6 provide an overview of the names, application examples and some important properties of the most common kinds of elastomers. Please note: The data specified there can serve only as inexact guideline values and must be verified in each individual case of an intended application.

Use of elastomers

Elastomers have the property of being able both to absorb large deformations reversibly and to absorb mechanical energy. Ultimately, this results in diverse

Table 5: Name and application examples of elastomers

Code	Designation	Application examples
M-group		
ACM	Acrylate rubber	Oil circuit (for example, O-rings, radial shaft seal)
AEM	Ethylene acrylate rubber	Oil circuit, dampers, absorbers
EPDM	Ethylene-propylene-diene rubber	Coolant circuit, brake parts, body seals
FFKM	Perfluorinated rubber	Little use (special applications)
FKM	Fluorcarbon rubber	Standard for fuel applications (gas and diesel)
O-group		
ECO	Epichlorhydrin rubber	Fuel hoses, membranes
Q-group		
FVMQ	Fluorosilicone rubber	Membranes, fuel applications
VMQ	Silicone rubber	Turbocharger hoses, exhaust suspension, airbag coating
R-group		
CR	Chloroprene rubber	Bellows, grommets, windshield wipers, V-belts
HNBR	Hydrogenated (acrylo)nitrile-butadiene rubber	Seals in the engine compartment, hoses, drive belts
IIR	Isobutene-isoprene rubber (butyl rubber)	Gas-tight elastomer parts (inner layer of tires), dampers, membranes
NBR	(Acrylo)nitrile-butadiene rubber	Seals, dampers, absorbers, membranes, valves
NR	Natural rubber	(Truck) tires, engine suspensions, chassis mounts
SBR	Styrene butadiene rubber	Passenger car tires, brake parts
U-group		
AU / EU	Polyurethane rubber (polyester / polyether)	Gear wheels, wipers, damping elements (such as foamed material), interior

Table 6: Properties of elastomers

Code	Hardness range in Shore A	Service temperature (continuous) in °C	Resistance to					
			Weather and ozone	Oils (mineral oil, engine oil)	Gasoline	Diesel fuel	Water	Brake fluid (glycol-based)
M-group								
ACM	50 to 90	–25 to +150 [1]	1–2	1	3–4	3	4	4
AEM	50 to 90	–35 to +150	1	1–2	3–4	3	3	4
EPDM	30 to 95	–50 to +125 [2] –50 to +150 [3]	1	4	4	4	1	1
FFKM	60 to 90	–15 to +260	1	1	1	1	1	1
FKM	55 to 90	–20 to +200 [1]	1	1	1	1	2	4
O-group								
ECO	50 to 90	–40 to +120	1–2	2	2	2	3	4
Q-group								
FVMQ	30 to 80	–55 to +175	1	1	2	1–2	1–2	4
VMQ	20 to 80	–60 to +200	1	2–3	4	3	1–2	2–3
R-group								
CR	30 to 90	–40 to +110	2–3	2–3	3–4	3	2–3	3
HNBR	40 to 90	–30 to +130 [4]	2	1	2	1	1	2
IIR	40 to 85	–40 to +120	2–3	4	4	4	1	1
NBR	35 to 95	–30 to +100 [4]	3–4 [6]	1	2	1	1–2	3–4
NR	30 to 95	–55 to +80	4	4	4	4	1	1
SBR	30 to 95	–50 to +100	4	4	4	4	1	1
U-group								
AU / EU	50 to 98	–40 to +90 [5]	2	1–2	3	2–3	4	4

1 Very good resistance (no or little effect)
2 Good resistance (moderate effect)
3 Limited resistance (significant effect)
4 No resistance (unsuitable)

[1] Special materials with better cold resistance are possible
[2] Sulfur cross-linked
[3] Peroxide cross-linked
[4] Cold resistance, depending on the composition of the polymer
[5] Materials with better resistance to heat and hydrolysis are possible
[6] Materials with better ozone resistance are possible

applications for this material class. Thus rubber compounds are used to manufacture products that bridge tolerances, permit movement between various components, constitute static and dynamic seals, diminish vibrations and act as springs.

Typical technical applications include tires, seals (see also Elastomer materials, Seals), V-belts, hoses, flexible couplings, cable jackets, bearing elements, retaining elements, shock absorbers, windshield wipers, conveyor belts, roofing foils and shoe soles. But elastomers are also used to manufacture items for everyday use such as rubber boots, erasers, rubber bands, balloons, condoms, rubber gloves, pacifiers and wet suits.

Thermosets

Thermosets are the first plastics to have been manufactured for industrial production using a synthetic method. As early as 1910, Bakelite, a resin made of formaldehyde and phenol, came into use. In the 1920s and 1930s, molding resin masses with urea and melamine jointed it on the market. Polyester and epoxy resins were used for the first time after the end of World War II [17].

In general, thermosets refer to plastics with close-meshed molecular chains that are tightly cross-linked to each other. This results in the typical properties of thermosets:

– High strengths and rigidities with simultaneously low density,
– High dimensional stability under heat and thermal resistance,
– Good resistance to chemicals,
– High brittleness.

The varied types of thermosets find application in a wide variety of areas in the automotive industry. In addition to structural components, which will be covered in more detail below, this includes paint and adhesive systems, sealing compounds and substrates for interconnect devices in electronics.

The good thermal properties of thermosets make them particularly suitable for use in thermally stressed areas. This specifically relates to applications in the engine compartment of automobiles. Typical examples of components made of thermoset include water pump housings, belt pulleys and impeller wheels [18]. New developments in the area of tribologically modified PF molding compounds also open up further use options, such as for bearing elements, sliding elements and guide elements [19].

Due to the large number of thermosetting resin systems, the section below only covers the technically relevant materials (Table 7).

Unsaturated polyester resins

Unsaturated polyester resins (UP resins) fully cure through radical polymerization under the influence of heat. They can be adjusted with almost no shrinkage, which leads to very good dimensional stability. This is why components with special requirements for precision, such as headlamp reflectors, are manufactured from these kinds of materials. Furthermore, components made of UP resins feature good electrical properties, which helped them find increased usage earlier in ignition systems of motor vehicles (ignition distributors).

Epoxy resins

Epoxy resins (EP resins) cure in a polymerization reaction and therefore, unlike the phenol-formaldehyde resins, do not split off any volatile reaction products. Compared to other thermosetting materials, EP resins have a particularly low viscosity, which simplifies the procedure for processing them. Due to the high material prices, however, use of epoxides is limited. EP resins are used, for example, for encapsulating electronic components (special low-pressure EP is used for this). Moreover, EP resin is increasingly being used in continuous-fiber-reinforced (mostly carbon fiber-reinforced) structural lightweight components as matrix materials in the automotive industry.

Table 7: Thermosets

Type	Type of resin	Filler	t_G [1] °C	σ_{bB} [2] min. N/mm²	α_n [3] min. kJ/m²	CTI [4] min. grade	Properties, application examples
Thermosets (new standards) – Pourable phenol molding compounds (PF–PMC) DIN EN ISO 14526 [33] – Pourable melamine-formaldehyde molding compounds (MF–PMC) DIN EN ISO 14528 [34] – Pourable melamine/phenolic molding compounds (MP–PMC) DIN EN ISO 14529 [35] – Pourable unsaturated polyester molding compounds (UP–PMC) DIN EN ISO 14530 [36] – Pourable epoxy-resin molding compounds (EP–PMC) DIN EN ISO 15252 [37]							
(WD30 + MD20) to (WD40 + MD10) (31 and 31.5) [7]	Phenol [8]	Wood flour	160/140	70	6	CTI 125	For parts subject to high electrical loads. [9]
(LF20 + MD25) to (LF30 + MD15) (51) [7]		Cellulose [5]	160/140	60	5	CTI 150	For parts with good insulating properties in low-voltage range. Type 74 impact-resistant. [9]
*SS40 to SS50 (74) [7]		Cotton fabric shreds [5]	160/140	60	12	CTI 150	
(LF20 + MD25) to (LF40 + MD05) (83) [7]		Cotton fibers [6]	160/140	60	5	CTI 150	Tougher than type 31. [9]
–		Glass fibers, short	220/180	200	12	CTI 125	High mechanical strength. Very good resistance to automotive fluids, low swelling.
–		Glass fibers, long	220/180	230	17	CTI 175	
–		Carbon fibers	220/180	250	14	–	High rigidity, low density, good wear properties. Not suitable for electrical applications (conductive).
(WD30 + MD15) to (WD40 + MD05) (150) [7]	Melamine	Wood flour	160/140	70	6	CTI 600	Resistant to glow heat, superior electrical properties, high shrinkage factor.

Table 7 (continued): Thermosets

Type	Type of resin	Filler	t_G [1] °C	σ_{bB} [2] min. N/mm²	α_n [3] min. kJ/m²	CTI [4] min. grade	Properties, application examples
Thermosets							
LD35 to LD45 (181) [7]	Melamine-phenol	Cellulose	160/140	80	7	CTI 250	For parts subject to electrical and mechanical loads.
(GF10 + MD60) to (GF20 + MD50) (802 and 804) [7]	Polyester	Glass fibers, inorganic fillers	220/170	55	4.5	CTI 600	Types 801, 804: low molding pressure required (large-area parts possible); types 803, 804: glow-heat resistant.
MD65 to MD75	Epoxy	Rock flour	200/170	80	5	CTI 600	Very good dielectrical properties. Sheathing sensors and actuators.
(GF25 + MD45) to (GF35 + MD35)		Glass fibers/ mineral	230/190	160	10	CTI 250	
–	Epoxy low-pressure molding compounds	SiO_2 (spherical)	250/200	120	6	–	Chip encapsulation (thin-bond wires).

[1] Maximum service temperature, short term (100 h)/continuous (20,000 h). [2] Flexural strength. [3] Impact strength (Charpy).
[4] Tracking resistance according to DIN IEC 112, Comparative Tracking Index (CTI).
[5] With or without addition of other organic fillers. [6] And/or wood flour. [7] Old designation in parentheses.
[8] Do not use types 13 to 83 (purely organically filled compounds) for new applications (availability no longer guaranteed).
[9] Barely used any more in new applications. Supply not guaranteed in the long term.

Phenol-formaldehyde resins
Phenol-formaldehyde resins (PF resins) are manufactured from phenol and formaldehyde using polycondensation. They have high mechanical rigidities and strengths and feature high chemical resistance, thermal resistance and dimensional stability under heat. In addition, there are tribologically modified forms that, however, cost significantly more because carbon fibers are added. PF resins are inherently flame retardant, which makes them ideal for applications in engine compartments.

Extenders
Filler are classified in accordance with DIN EN ISO 1043-2 [20]. A distinction is made here between material (such as carbon, glass, mineral), the form and structure (such as fibers, balls, powder) and special properties (such as flame retardant, thermal resistance) [4]. Technically relevant fillers primarily include glass fibers and glass balls as well as mineral fillers. Particularly in the aerospace industry, but also increasingly in the automotive industry, carbon fibers are also being used as reinforcement materials for thermosets.

Processing method
Thermosetting molding compounds are primarily processed using compacting or injection molding. There is also a wide variety of other processes in which the listed variants are combined or further developed.

Compacting
For compacting, pressure and temperature simultaneously act upon the molding compound. The mass is fed into the mold either without a form or in tablet form. Pressure brings the mass into the desired form, while the temperature cross-links the material. Due to the low machine-based effort, compacting is the least expensive processing method. However, this method can be used to manufacture only simple, large-format components. Above all, the manufactured components feature low level of orientation and the need for large amounts of finishing.

Injection molding
Here it is necessary to distinguish between easy-flowing granulate material and polyester bulk molding compound. Exceptionally high productivity is achieved when processing easy-flowing granulate material where the length of cycle times is largely dependent on component thickness. The barrel temperatures during injection molding are between 80 and 100 °C, while the mold temperatures range from 160 to 190 °C. While the injected mass is curing in the mold, new material for the next cycle is already being plasticized in the screw.

When processing polyester bulk molding compound (UP BMC), it is important to note that a stuffing unit is required in addition to the conventional machine equipment to convey the material into the screw. Then the material is homogenized in the injection barrel. Melting is not necessary; the barrel temperatures are usually around 25 to 35 °C.

Insulating materials
Electrical insulation plays a critical role in the proper functioning and service life of items such as alternators, engines and electrical devices, not just in motor vehicles.

Unfilled polymers have the best electrical insulation properties. Each addition of fillers reduces the dielectric strength of a polymer material by forming interfaces between the filler and polymer matrix as well as by excessive stress due to different dielectric properties.

The electrical insulation in motor vehicles not only has to reliably prevent disruptive electrical breakdowns, it also has to dissipate heat losses that arise, absorb mechanical loads and be resistant to liquids commonly used in motor vehicles. These properties typically have to be ensured in a temperature range from –40 to +180 °C. For this reason, a system of several materials is usually used as an insulating system; a single material is rarely used.

As an example here we will discuss surface insulators. These are usually used as more or less flexible combined insulation materials in electrical machines such as starters, alternators, hybrid engines and traction motors.

Plastic foils are flexible combined insulation materials that are bonded to pressboard, nonwoven materials or paper made of polymers. Flexible combined insulation materials often have a three-ply design with plastic foil as the middle layer. Depending on the combination of materials, they have different continuous service temperatures, tensile strengths and elongations at break, dielectric strengths, rigidities and impregnabilities.

The combination of plastic foils with fiber or nonwoven materials has technical advantages. Plastic foil made of unfilled polymers provides excellent electrical properties. The fiber or nonwoven materials, by contrast, have good impregnability and protect the foil against mechanical and thermal loads.

Foils made of polyester or polyimide are predominantly used as initial components as well as fiber or nonwoven materials made of organic fibers, polyester or aramide.

Flexible combined insulation materials are standardized as DIN EN 60626-1 ([21], definitions, general requirements), DIN EN 60626-2 ([21], test methods) and DIN EN 60626-3 ([21], properties of individual combinations of materials). In addition, there are standards for other surface insulators.

- Insulating foils: DIN EN 60674-1 ([22], definitions, general requirements), DIN EN 60674-2 ([22], test methods), DIN EN 60674-3-1 to -3-8 ([23], [24], properties of individual materials).
- Panel and roller pressboard: DIN EN 60641-1 ([25], definitions, general requirements), DIN EN 60641-2 ([26], test methods), DIN EN 60641-3-1, -3-2 ([26], properties of individual materials).
- Laminates: DIN EN 60893-1 ([27], definitions, general requirements), DIN EN 60893-2 ([27], test methods), DIN EN 60893-3-1 to -3-7 ([27], properties).

Sealing compounds

Sealing compounds (or cast resins) refer to reactive synthetic resins that are processed as a liquid for the finished product and solidify as this or a component part of it. While still a liquid, the resin is poured into a reusable or disposable mold. This results in either pure cast resin bodies with free form surfaces or other parts are included ([28], [29], [30], [31], [32]).

Unlike with meltable plastics (thermoplastics), the solidification occurs through a chemical cross-linking reaction and is irreversible (thermoset). They are usually poured in

- To cover and protect parts against penetration of moisture, dust, foreign matter, water, etc.
- To fix parts into place with respect to each other
- To increase the mechanical stability and the vibration and shock resistance
- To provide electrical insulation, that is, to increase the dielectric strength and contact protection
- To dissipate any heat losses that arise.

The range of applications for sealing compounds is broad and multi-faceted. As a result, the functions and requirements are also multi-faceted, ranging from processing and curing to the properties of the later application area. Selection criteria include:

- For application of liquid, not yet hardened sealing compound: viscosity, pot life, available systems engineering (manual, metering or molding system).
- The requirements that are placed on the cured molding compound in later use, such as hardness, elasticity, stretchability or pliability, thermal, mechanical, electrical and chemical loads of the molded components.

Due to their mechanical and electrical properties (Table 8), sealing compounds are ideal for use in electrical engineering and electronics. Typical applications of sealing compounds include:

– Molding and manufacturing electrotechnical components (such as ignition coils, transformers, insulators, capacitors, semiconductors, assemblies)
– Molding of open contact points for cables and lines

Different material classes are used depending on the requirements profile (such as service temperatures, chemical resistance, mold geometry). The sealing compounds used most frequently in electronics potting belong to the epoxide, polyurethane and silicone material classes. Less widespread, for example, are polyester and hot melt sealing compounds (hot melts, hot melt adhesives).

Table 8: Properties of different sealing compounds (examples)

Properties	Standard	Unit	Epoxy resin, unfilled	Epoxy resin, filled (quartz powder)	Unsaturated polyester resins	Polyurethanes	Silicone
Tensile strength	ISO 527-1 [38]	N/mm^2	60 to 90	80 to 100	30 to 80	3 to 80	0.3 to 10
Elongation at break	ISO 527-1 [38]	%	3 to 8	0.8 to 1.1	2 to 4	0.5 to 80	50 to 700
Flexural strength	ISO 178 [39]	N/mm^2	60 to 140	110 to 120	60 to 140	40 to 140	n/a
Impact strength	ISO 179-1 [40]	kJ/m^2	15 to 30	10 to 12	8 to 15	40	n/a
Glass transition temperature T_G		°C	+70 to +200	+70 to +200	+70 to +150	−40 to +130	−120
E module		N/mm^2	2,000 to 4,000	5,000 to 8,000	3,500	250 to 3,000	0.005 to 5
Continuous temperature resistance	DIN EN 60216-1 [41]	°C	110 to 200	110 to 200	120 to 140	90 to 140	150 to 250
Specific volume resistance		Ω cm	10^{14} to 10^{16}	10^{14} to 10^{16}	10^{13}	10^{13} to 10^{16}	10^{15} to 10^{17}
Exothermy during curing			High	Medium	Very high	Very low	Very low
Coefficient of thermal expansion		$K^{-1}\ 10^{-6}$	80 to 100	30 to 70	60 to 80	50 to 150	300
Shrinkage		%	0.5 to 2	0.1 to 1	3 to 9	0.2 to 1	0.1 to 2

+ = Good resistance
O = Limited resistance
– = No resistance
n/a = Not applicable

There are also extensive means of using chemicals and formulas to adapt the sealing compounds to the typical requirements for an application:
– Variations in chemicals provide different resin and curing systems
– Variations in formulas provide fillers, pigments, accelerators, toughness modifiers, wetting agents, degassing agents and anti-sedimentation agents

The chemical resistance of sealing compounds depends on the components of the cast resin, the cross-link density and the degree of cross-linking. As a rule of thumb, hard sealing compounds are more resistant than soft ones.

The final properties of the cured sealing compounds largely depend on the kind and quantity of ingredients used (hardeners, diluents, fillers) and the curing conditions. Therefore generally applicable specifications about physical properties are not possible.

Epoxide systems
Epoxides have been widely used for many years. They are generally hard and capable of carrying loads, their volume shrinkage during curing is rather low.

Characteristics include their excellent mechanical properties, good temperature tolerance, good adhesive strength on a wide variety of surfaces, and likewise good chemical resistance. The cross-linking or curing process is generally slow, particularly if only small volumes react with each other. More reactive hardeners could be used, but that can cause a substantially stronger exothermic reaction, creating stress for components and the printed circuit board.

Fillers are an essential component of epoxy resin formulations. They reduce costs and improve mechanical properties. At the same time, they lead to an increase in the crack resistance and rigidity, as well as to a reduction in thermal shrinkage, lowering lower inner stresses. Quartz powder is a standard filler in electronic applications.

Epoxide systems are frequently used, for example, as a thermally conductive sealing compound for all kinds of solenoids.

Polyurethane systems
Even after curing, polyurethane sealing compounds can still expand and flex (even if only somewhat), which is particularly important when sensitive components (such as ferrite on printed circuit boards) are cast.

The curing reaction for polyurethane systems is less exothermic than for epoxide sealing compounds. The volume shrinkage after the hardening is low and a broad spectrum of hardness and/or elasticity is available. The chemical and mechanical resistances are good.

Silicone
Silicone-based sealing compounds are usually significantly more expensive than epoxide or polyurethane sealing compounds, but they are used where typical continuous service temperatures are above 180 °C. The exothermic heat buildup during the hardening is also only very low.

Polyester systems
These systems have very good resistances to a wide variety of media, but a very strong heat buildup during hardening and a high degree of shrinkage after the hardening as a result. Under certain circumstances this leads to thermomechanical damage to the cast components, even to the point of having components tear away from the substrate.

References for plastics
[1] M. Gehde, S. Englich, G. Hülder, M. Höer: Schlummerndes Potenzial für den Leichtbau [The dormant potential for lightweight construction]. In: Plastverarbeiter (10), p. 80-83, 2012.
http://www.plastverarbeiter.de/36471/schlummerndes-potenzial-fuer-den-leichtbau/
[2] P. Elsner, P. Eyerer, T. Hirth: Domininghaus – Kunststoffe: Eigenschaften und Anwendungen [Plastics: Properties and applications]. 8. Aufl., Springer, 2012.
[3] DIN 7724:1993: Polymere Werkstoffe – Gruppierung polymerer Werkstoffe aufgrund ihres mechanischen Verhaltens [Polymer materials – Grouping polymers by their mechanical characteristics].
[4] W. Hellerich, G. Harsch, S. Haenle: Werkstoff-Führer Kunststoffe – Eigenschaften, Prüfungen, Kennwerte [Plastics material guide – Properties, Testing and Characteristics]. 10. Aufl., Carl Hanser Verlag, 2010.
[5] E. Baur, S. Brinkmann, N. Rudolph, T.A. Osswald, E. Schmachtenberg: Saechtling Kunststoff Taschenbuch [Saechtling plastics pocketbook], 31. Aufl., Carl Hanser Verlag, 2013.
[6] G.W. Ehrenstein, S. Pongratz: Beständigkeit von Kunststoffen [The durability of plastics]. 1. Aufl., Carl Hanser Verlag, 2007.
[7] DIN 50035:2012: Begriffe auf dem Gebiet der Alterung von Materialien – Polymere Werkstoffe [Concepts in the field of material aging – Polymers].
[8] T. Dolansky, M. Gehringer, H. Neumeier: TPE-Fibel – Grundlagen, Spritzguss [TPE primer – Principles, casting]. 1. Aufl., Dr. Gupta Verlag, 2007.
[9] F. Röthemeyer, F. Sommer: Kautschuk Technologie: Werkstoffe – Verarbeitung – Produkte [Rubber technology: Materials – Processing – Products]. 3. Aufl., Carl Hanser Verlag, 2013.
[10] G. Holden, H.R. Kricheldorf, R. Quirk: Thermoplastic Elastomers [Thermoplastic elastomers]. 3. Aufl., Hanser Gardner Publ., 2004.
[11] DIN EN ISO 18064: Thermoplastic elastomers – Nomenclature and abbreviated terms.
[12] K. Nagdi: Gummi-Werkstoffe [Rubber materials]. 3. Aufl., Dr. Gupta Verlag, 2004.
[13] W. Hofmann, H. Gupta: Handbuch der Kautschuk-Technologie [Handbook of rubber technology]. Dr. Gupta Verlag, 2001.
[14] J. Schnetger: Lexikon Kautschuktechnik [Glossary of rubber technology]. 3. Aufl., Beuth, 2004.
[15] G. Abts: Einführung in die Kautschuktechnologie [Introduction to rubber technology]. Carl Hanser Verlag, 2007.
[16] DIN ISO 1629: Rubber and latices – Nomenclature.
[17] G.W. Becker, D. Braun, W. Woebcken (Hrsg.): Kunststoff Handbuch – Band 10: Duroplaste [Handbook of Plastics – Volume 10: Thermosets]. 2. Aufl., Hanser Fachbuch, 1988.
[18] E. Bittmann: Duroplaste kommen ins Rollen [Thermosets get rolling]. In: Kunststoffe 3/2003, A25-A27.
https://www.kunststoffe.de/kunststoffe-zeitschrift/archiv/artikel/automobilanwendungenduroplaste-kommen-ins-rollen-530048.html.
[19] E. Bittmann: Duroplaste [Thermosets]. In: Kunststoffe 10/2005, 168–172.
https://www.kunststoffe.de/kunststoffe-zeitschrift/archiv/artikel/hoffnungsvolle-auftragslage-duroplaste-und-fvk-537533.html.
[20] DIN EN ISO 1043-2: Plastics – Symbols and abbreviated terms – Part 2: Fillers and reinforcing materials. [21] DIN EN 60626: Combined flexible materials for electrical insulation. Part 1: Definitions and general requirements.
Part 2: Methods of test.
Part 3: Specifications for individual materials.
[22] DIN EN 60674: Plastic films for electrical purposes. Part 1: Definitions and general requirements.
Part 2: Methods of test.
[23] DIN EN 60674-3-1: Plastic films for electrical purposes – Part 3: Specifications for individual materials.
Sheet 1: Biaxially oriented polypropylene (PP) film for capacitors.
[24] DIN EN 60674-3-2: Specification for plastic films for electrical purposes – Part 3: Specifications for individual materials.
Sheet 2: Requirements for balanced biaxially oriented polyethylene terephthalate (PET) films used for electrical insulation.

Sheet 3: Requirements for polycarbonate (PC) films used for electrical insulation.
Sheets 4 to 6: Requirements for polyimide films used for electrical insulation.
Sheet 7: Requirements for fluoroethylene-propylene (FEP) films used for electrical insulation.
Sheet 8: Balanced biaxially oriented polyethylene naphthalate (PEN) films used for electrical insulation.
[25] DIN EN 60641-1: Specification for pressboard and presspaper for electrical purposes.
Part 1: Definitions and general requirements.
[26] DIN EN 60641-2: Pressboard and presspaper for electrical purposes.
Part 2: Methods of test.
Part 3: Specifications for individual materials.
Sheet 1: Requirements for pressboard.
Sheet 2: Requirements for presspaper.
[27] DIN EN 60893-1: Insulating materials – Industrial rigid laminated sheets based on thermosetting resins for electrical purposes.
Part 1: Definitions, designations and general requirements.
Part 2: Methods of test.
Part 3-1: Specifications for individual materials – Types of industrial rigid laminated sheets.
Part 3-2: Specifications for individual materials – Requirements for rigid laminated sheets based on epoxy resins.
Part 3-3: Specifications for individual materials – Requirements for rigid laminated sheets based on melamine resins.
Part 3-4: Specifications for individual materials – Requirements for rigid laminated sheets based on phenol resins.
Part 3-5: Specifications for individual materials – Requirements for rigid laminated sheets based on polyester resins.
Part 3-6: Specifications for individual materials – Requirements for rigid laminated sheets based on silicone resins.
Part 3-7: Specifications for individual materials – Requirements for rigid laminated sheets based on polyimide resins.
[28] R. Stierli: Epoxid-Gieß- und Imprägnierharze für die Elektroindustrie [Expoxide casting and impregnating resins for the electrical industry]. In: Wilbrand Woebcken (Hrsg.): Duroplaste – Kunststoff-Handbuch [Thermosets – The plastics handbook], Band 10. 2. Aufl., Hanser Fachbuch, 1988.
[29] W. Becker, D. Braun, G. Oertel (Hrsg.): Polyurethane – Kunststoff-Handbuch [Polyurethane – The plastics handbook], Band 7. 3. Aufl., Hanser Fachbuch, 1993.
[30] Dr. Werner Hollstein, Huntsman Advanced Materials GmbH: Einführung in die Chemie der Epoxidharze und Formulierungskomponenten [Introduction to the chemistry of epoxide resins and formulation components].
[31] http://www.electrolube.com.
[32] Lackwerke Peters GmbH & Co. KG: Technische Informationen TI 15/2.
https://www.peters.de/de/download-center.
[33] DIN EN ISO 14526-1: Plastics – Phenolic power moulding compounds (PF-PMCs) – Part 1: Designation system and basis for specifications.
[34] DIN EN ISO 14528-1: Plastics – Melamine-formaldehyde powder moulding compounds (MF-PMCs) – Part 1: Designation system and basis for specifications.
[35] DIN EN ISO 14529-1: Plastics – Melamine/phenolic power moulding compounds (MP-PMCs) – Part 1: Designation system and basis for specifications.
[36] DIN EN ISO 14530-1: Plastics – Unsaturated-polyester power moulding compounds (UP-PMCs) – Part 1: Designation system and basis for specifications.
[37] DIN EN ISO 15252-1: Plastics – Epoxy power moulding compounds (EP-PMCs) – Part 1: Designation system and basis for specifications.
[38] ISO 527-1: Plastics – Determination of tensile properties – Part 1: General principles.
[39] ISO 178: Plastics – Determination of flexural properties.
[40] ISO 179-1: Plastics – Determination of Charpy impact properties – Part 1: Non-instrumented impact test.
[41] DIN EN 60216-1: Electrical insulating materials – Thermal endurance properties – Part 1: Ageing procedures and evaluation of test results.

Heat treatment of metallic materials

Hardness

Hardness is a property of solid materials that defines the resistance of a materials against penetration by a harder solid body. In metallic materials the hardness is used to assess mechanical properties such as strength, machinability, malleability, and resistance to wear. DIN EN ISO 18265 [1] defines guidelines for converting hardness to tensile strength.

Hardness testing

Hardness testing is a nondestructive way of obtaining information about the mechanical properties of a material in a relatively short time.

The test data are generally derived from the size or depth of the deformation produced when a specified indentor is applied at a defined pressure. A distinction is made between static and dynamic testing. Static testing is based on measurement of the permanent impression left by the indentor. Conventional hardness tests include the Rockwell, Vickers and Brinell procedures. Figure 1 compares the fields of application of hardness testing based on these different procedures.

Dynamic testing monitors the rebound height of a test tool accelerated against the surface of the test specimen.

Another option for obtaining an index of surface hardness is to scratch the surface with a harder test tool and then measure the groove width.

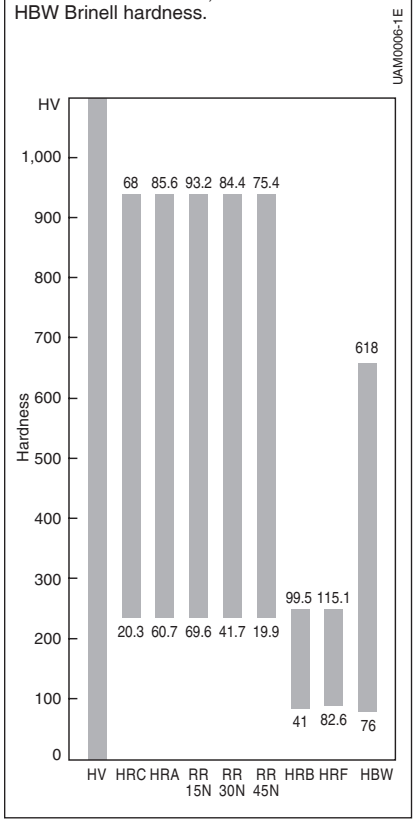
Hardness-testing methods

Rockwell hardness (DIN EN ISO 6508, [2])
This method is particularly suitable for fast, automated testing of metallic workpieces, but places specific demands on clamping the workpiece in the test equipment. It is unsuitable for workpieces which as a result of their geometry give in the test equipment (e.g. pipes).

This method involves a test tool (indentor) of defined size, shape (tapered or spherical) and material (steel, hard metal or diamond) being pressed in two stages into the test specimen. In this process, after the preliminary test force is applied, the additional test force is exerted for a defined period. Initially, the factor e is calculated from the remaining penetration depth h for the HRA, HRB, HRC and HRF methods:

$$e = \frac{h}{0.002}.$$

Figure 1: Comparison of hardness ranges for different test methods for unalloyed and low-alloy steels and cast steel
The figures at the range extremities indicate the hardness data for the respective method.
HV Vickers hardness,
HR Rockwell hardness,
HBW Brinell hardness.



For the HR..N and HR..T methods the factor e is calculated from the penetration depth h with the equation:

$$e = \frac{h}{0.001}.$$

With the factor e the Rockwell hardness HR can now be calculated for the HRA, HRC, HR..N and HR..T methods:

$$HR = 100 - e.$$

If the HRB and HRF methods are used, the Rockwell hardness HR is calculated according to the of following equation:

$$HR = 130 - e.$$

The test specimen's surface should be smooth and as flat as possible. When testing on convex-cylindrical or spherical surfaces, the value determined must be corrected as a function of the hardness.

The abbreviation for the selected test method should be appended to the numerical value when specifying hardness (e.g.: 65 HRC, 76 HR45N). The designations provide an indication of the indentor used (diamond cone or ball), the preliminary test force and the total test force.

Table 1: Application of the Brinell hardness test

Material	Brinell hardness	Load factor $0.102\ F/D^2$
Steel; nickel and titanium alloys		30
Cast iron (nominal diameter of ball must be 2.5, 5 or 10 mm)	< 140 ≥ 140	10 30
Copper and copper alloys	< 35	5
	35 to 200	10
	>200	30
Light metals and their alloys	< 35	2.5
	35 to 80	5 10 15
	> 80	10 15
Lead and tin		1
Sintered metals	see DIN EN ISO 4498 [4]	

Depending on the indentor used and the total test force applied, different hardness scales with the abbreviations HRA, HRB, HRC, HRD, HRE, HRF, HRG, HRH, HRK, HR15N, HR30N, HR45N, HR15T, HR30T and HR45T are used.

Advantages of the Rockwell test method include minimal specimen preparation and rapid measurement. This test process can also be fully automated. Any tester vibration, or shifting and movement of either the test specimen or the indentor can lead to testing errors, as can an uneven support surface or a damaged indentor.

Brinell hardness (DIN EN ISO 6506, [3])
This method is used for metallic materials of low to medium hardness. The test tool (indentor) is a hard-metal ball with the diameter D. It is pressed with test force F vertically into the surface of a specimen. After the test force has been removed, the Brinell hardness is calculated from the remaining indentation diameter d:

$$HBW = 0.102\ \frac{2F}{\pi D^2 (1 - \sqrt{1 - d^2/D^2})},$$

with load F in N, ball diameter D in mm, and mean indentation diameter d in mm.

Test loads range from 9.81 N to 29,420 N. Results obtained using balls of different diameters are only conditionally comparable, and any comparisons should be based on testing at identical force levels. Testing should always be performed using the largest possible ball, while load factors should be selected to obtain an indentation diameter of between 0.24 D and 0.6 D. Table 1 lists the recommended load factors and ball diameters for a variety of materials as laid down in DIN EN ISO 6506-1 [3].

In the Brinell-hardness designation, the numeric value is accompanied by the abbreviation for the method, the ball diameter in mm and the test force in N multiplied by a factor of 0.102 (e.g. 600 HBW 1/30).

High test loads producing deformations extending over a relatively wide surface area can be employed to gather data on materials with inconsistent structures. An advantage of the Brinell method is the relatively high degree of correlation between the Brinell hardness factor and the steel's tensile strength.

The required preparations and ensuing test procedures are more complex than those used for Rockwell testing.

Vickers hardness (DIN EN ISO 6507, [5])
This test method can be used for all metallic materials, regardless of hardness. It is especially suitable for testing minute and thin specimens, while the potential application range extends to include surface and case-hardened parts, as well as nitrided workpieces and parts carburized in nitrogen-based atmospheres.

The test tool is a square-based diamond pyramid with an apical angle of 136°. This is applied to the surface of the test specimen at an individually specified force F. The diagonals d_1 and d_2 of the indentation remaining on the test surface after test force F has been removed are measured. The Vickers hardness is obtained from the arithmetic mean value d of the two diagonals:

$$\text{HV} = \frac{2F \sin\frac{136}{2}}{d^2} \approx 0.1891 \frac{F}{d^2},$$

with test force F in N and arithmetic mean value d of diagonal lengths d_1 and d_2 in mm.

In the formula for Vickers hardness data, the actual test figure is accompanied by the abbreviation HV, the force in N (multiplied by a factor of 0.102) and, following a slash, the force application period (if other than the standard 15 seconds) in seconds (e.g.: 750 HV 10/25).

The test specimen's surface should be smooth and flat. DIN EN ISO 6507 [5] stipulates that correction factors be used to compensate for any error stemming from surface curvature. Test-load levels are selected with reference to either the thickness of the workpiece to be tested or the layer to be tested.

A major advantage of this test method is that there are virtually no limitations on using it to assess thin parts or layers. It allows the use of extremely small force levels to determine the hardness of individual structural sections. The Brinell and Vickers numbers correlate well up to approximately 350 HV. However, a certain minimal degree of surface consistency is necessary to ensure accurate results.

Knoop hardness (DIN EN ISO 4545, [6])
This process closely resembles the Vickers method. The equal-sided diamond tip in the Vickers test has a rhombic shape in the Knoop test. The test tool is designed to leave an impression in the form of a thin, elongated rhombus. The long diagonal d of the indentation remaining on the test surface after test force F has been removed is measured. The hardness value HK is calculated as follows:

$$\text{HK} = 1.451 \frac{F}{d^2},$$

with test force F in N and length d of the long diagonal in mm.

In the Knoop-hardness designation, the numeric value is accompanied by the abbreviation HK for the method, the test force in N multiplied by a factor of 0.102 and – if necessary separated by a slash – the application period of the test force in s (e.g. 640 HK 0.1/20).

The penetration depth is roughly 1/3 less than in the Vickers method, allowing evaluation of surface hardness in thin parts and layers. However, the test method also places heightened demands on the surface: The test must be performed on a polished, smooth and flat surface.

Knoop testing is often used on brittle materials such as, for example, ceramic or sintered materials.

Martens hardness (DIN EN ISO 14577, [7])
In this method, the penetration of a pyramid-shaped test tool into materials is recorded, in the course of which both the force and the travel during plastic and elastic deformation are measured. Martens hardness is defined as the ratio of test force F to surface A_S of the indentor calculated from penetration depth h, and given in N/mm^2.

Martens hardness is denoted by the abbreviation HM, to which is appended the test force in N, the application time of the test force in s and the holding time of the test force in s (e.g. HM 0.5/20/20 ≙ 8,700 N/mm^2).

Scleroscope hardness
This dynamic measurement method is specially designed for heavy and large metal pieces. This process is based on the measurement of the rebound height of a steel indentor (hammer) featuring a diamond or hard metal tip. This is dropped from a stipulated height onto the surface of the test specimen. The rebound serves as the basis for determining the hardness.

The method is not standardized and there is no direct correlative relationship with any other hardness-testing method.

Heat-treatment processes

Heat treatment is employed to adapt the technological material properties of metallic components and tools to the relevant requirements. Such requirements can be processing properties to suit manufacturing and usage properties to suit function.

According to DIN EN 10052 [8], heat treatment involves "subjecting a workpiece in whole or in part to time and temperature cycles in order to bring about a change in its properties of its structure. If necessary, the chemical composition of the material can be changed during the treatment".

The process modifies the microstructure to achieve the hardness, strength, ductility, wear resistance, etc. required to withstand the stresses associated with static and dynamic loads. The most significant industrial processes are summarized in Table 2 (see DIN EN 10052, [8] for terminology).

Hardening
Hardening procedures produce a martensitic microstructure of extreme hardness and strength in ferrous materials such as steel and cast iron. These procedures consist of the individual stages known as austenitizing and cooling or quenching.

Through hardening
The workpiece being treated is heated to the austenitizing, or hardening temperature (Table 3), at which it is maintained until an austenitic structure emerges, and until an adequate quantity of carbon (released in the decay of carbides, such as graphite in cast iron) is dissolved in the treated material. After austenitizing, the workpiece is cooled or quenched at a rate sufficient for hardening, which can also occur in temperature stages. A complete conversion to the martensitic microstructure should be completed as quickly as possible. The necessary cooling process is determined by the chemical composition, the austenitizing conditions, the shape and dimensions, and the desired microstructure. Reference values for the required cooling rate can be found in the time-temperature transformation chart for the steel in question.

Table 2: Overview of heat-treatment processes

Hardening	Austempering	Draw tempering	Thermo-chemical treatment	Annealing	Precipitation hardening
Through hardening cross section Surface hardening Hardening of carburized parts (case-hardening)	Isothermic transformation in the bainite stage	Tempering of hardened parts Tempering above 540 °C for quenching and drawing	Carburizing Carbonitriding Nitriding Nitro-carburizing Boron treatment Chromating	Stress-free annealing Recrystallization annealing Soft annealing Spheroidization Normalizing Homogenizing	Solution treatment and aging

The austenitizing temperature varies according to the composition of the material in question (for specific data, consult the DIN Technical Requirements for Steels). Table 3 above provides reference data. See DIN 17022 [9] Part 1 and Part 2 for practical information on hardening procedures for tools and components.

Not all types of steel and cast iron are suitable for hardening. The following equation describes the hardening potential for alloyed and unalloyed steels with mass carbon contents of between 0.15 % and 0.60 %, and can be applied to estimate the hardness levels achievable with a completely martensitic microstructure:

Table 3: Standard austenitizing temperatures

Type of steel	Quality specification [10, 11, 12]	Austenitizing temperature in °C
Unalloyed and low-alloy steels	DIN EN 10083-1 10083-2 10083-3	
< 0.8 % by mass of C	10085	780...950
≥ 0.8 % by mass of C	–	750...820
Cold- and hot-working tool steels	DIN EN ISO 4957	780 to 1150
High-speed tool steels		1150 to 1300

$$\text{Max. hardness} = (35 + 50x \pm 2)\,\text{HRC}, \text{ (eq. 1)}$$

where the carbon content in % by mass is to be applied for x here. If the microstructure does not consist entirely of martensite, then the maximum hardness will not be achieved.

When the carbon content exceeds 0.6 % by mass, it may be assumed that the material's structure contains untransformed austenite (residual austenite) in addition to the martensite. Greater amounts of residual austenite can have a deleterious effect on achievable hardness and reduce wear resistance. In addition, residual austenite is metastable, i.e. there exists a potential for subsequent transformation to martensite at temperatures below room temperature or under stress. This can give rise to unwanted changes in the workpiece's dimensions and shape. Prompt low-temperature follow-up procedures or draw-tempering operations at over 230 °C can be useful in cases where residual austenite is an unavoidable product of the hardening procedure.

Temperature differences occur between the edge and core of the workpieces during cooling and quenching. With greater cross-sections, the reduced cooling rate in the core can give rise to a decrease in hardness as the distance to the surface increases. There is a hardness progression or gradient. The hardness

gradient is obtained from the material composition and the hardenability dependent on the austenitizing conditions (testing described in DIN EN ISO 642, [13]). In this case, it is necessary with regard to the required hardness to use material with sufficient hardenability. Information on choice of steel based on hardenability can be found in DIN 17021 [14].

DIN EN ISO 18265 [15] defines the method for using hardness as the basis for estimating tensile strength R_m. This method can only be applied in cases whether the surface and core hardnesses are virtually identical.

During hardening, the transformation of the microstructure to the martensitic state is combined with an increase in volume. With regard to the initial state, the volume increases by approximately 1 %. This equates to a change in length of approximately 0.3 %.

The changes in volume associated with rearrangements of the structure and thermal gradients during cooling give rise to stresses, which in turn result in distortion in the form of changes in dimension and shape. The stresses that remain in the workpiece after hardening are called internal stresses. The edge of a hardened workpiece tends to be subject to internal tensile stress, while the core tends to be subject to internal compressive stress.

Table 4: Comparison of power densities when heating with different sources

Energy source	Normal power density in W/cm^2
Laser beam	10^3 to 10^4
Electron beam	10^3 to 10^4
Induction (with medium- or high-frequency alternating current, or high-frequency pulses)	10^3 to 10^4
Flame heating	$1 \cdot 10^3$ to $6 \cdot 10^3$
Plasma beam	10^4
Molten saline solution (convection)	20
Air, gas (convection)	0.5

Surface hardening

This process is especially suited for integration within large-scale manufacturing operations, and can be adapted to fit the rhythm of the production line.

Heating and hardening are restricted to the surface, thereby minimizing alterations in shape and dimensions. Heating is generally provided by high- or medium-frequency alternating current (induction hardening) or by a gas burner (flame hardening). Friction (friction hardening) and high-energy beams (e.g. electron or laser beams) can also provide the heat required for austenitizing. Table 4 provides an overview of the specific heat energies for the individual procedures.

These methods can be used to treat both linear and flat surfaces, meaning that the parts can be heated either while stationary or in motion. The heat source itself can also be moved. Rotation is the best way of dealing with radially symmetrical parts, as it ensures concentric hardening. Either immersion or spraying arrangements can be applied for quenching. Information on performing surface hardening can be found in DIN 17022-5 [9].

Heat rise is rapid, so the temperatures must be 50 to 100 °C higher than those used in furnace heating so as to compensate for the shorter dwell period. The procedure is generally employed with low-alloy or unalloyed steels with mass carbon contents of 0.35 to 0.60 %. However, surface hardening processes can also be applied with alloyed steels, cast iron and rolling-bearing steels. The parts can be heat-treated to provide a combination of improved base strength and high surface hardness, making them suitable for high-stress applications (recessed edges, bearing surfaces, cross-sectional transitions).

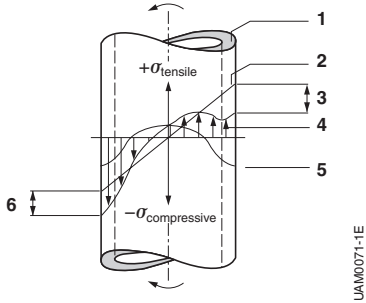
Surface hardening generally results in internal compression stresses along the edge. This leads to increased fatigue resistance, especially when notched parts are exposed to inconstant vibration stress. The stress in Figure 2 corresponds to bending stress. The higher stressability results from the fact that the stress state (resulting stress) is reduced from the superposition of bending stress and internal stress.

The relationship defined in equation 1 can be employed to estimate the potential surface hardness. There is a substantial reduction in hardness between the surface and the unhardened core region. The surface hardness depth SHD – the depth at which 80 % of the minimum Vickers surface hardness is found – can be derived from the hardness progression curve (see DIN EN 10328, [16]).

Austempering
The object of this process is to achieve a bainite microstructure. This microstructure is not as hard as martensite, but does display greater ductility, as well as smaller changes in specific volume.

After austenitizing (see hardening), the parts for austempering are first cooled to a temperature of 200 to 350 °C (depending upon the exact composition of the material) at the required rate. The parts are then held at this temperature until the microstructure's transformation into bainite has been completed. Cooling to the transformation temperature is usually performed in molten saline solutions (a typical salt is a mixture of potassium nitrate and sodium nitrite). Following the microstructure transformation the parts can be cooled to room temperature (no special procedure required).

Figure 2: Cyclically alternating stress according to surface-layer hardening
$+\sigma$ Tensile, $-\sigma$ Compressive.
1 Case layer, 2 Bending stress,
3 Reduction of tensile stress,
4 Resulting stress, 5 Internal stress,
6 Increase in compressive stress.



Austempering is an excellent alternative for parts whose geometrical configuration makes them sensitive to distortion or cracks, or in which high ductility is required together with substantial hardness, or which should combine hardness with a low level of residual austenite.

Application
Cylinder heads in modern diesel high-pressure pumps for common-rail systems which must be able to withstand high wear and internal compressive stresses are austempered.

Draw tempering
Draw tempering of hardened components and tools is employed to increase their deformability and to reduce their risk of cracking. According to DIN EN 10052 [8], draw tempering involves heating the part in question once or repeatedly to tempering temperature, holding the part at this temperature, and then allowing it to cool appropriately. Draw tempering is performed between room temperature and Ac_1 temperature, i.e. the temperature at which austenitic structural constituents would be created.

Tempering at temperatures as low as 180 °C is enough to reduce the hardness of unalloyed and low-alloy steels by approx. 1 to 5 HRC. The individual materials respond to higher temperatures with specific characteristic hardness loss. Figure 3 shows a characteristic tempering curve for typical types of steel. This graph illustrates the fact that the hardness of steels which are alloyed with special carbide-forming elements (Mo, V, W) – such as, for example, hot-working or high-speed tool steel – is increased by tempering in a temperature range between 400 and 600 °C to values which can be above the quenching hardness (secondary hardening).

The mutual relationships between tempering temperature on the one side, and hardness, strength, yield point, fracture contraction and elongation at fracture on the other, can be taken from the tempering diagrams for the various steels (see e.g. DIN EN 17021, [17]).

Generally speaking, draw tempering reduces hardness and strength and increases deformability. Internal stresses

can also be reduced at tempering temperatures in excess of 300 °C.

The specific volume decreases when structures which are free of residual austenite are tempered. In the case of structures containing residual austenite, however, an increase in volume occurs during the transformation from residual austenite to martensite. Hardness increases, deformability is reduced, and new internal stresses can be created. The risk of cracking also increases.

It must be remembered that steels alloyed with manganese, chromium and nickel, or combinations of these elements, should not be tempered at temperatures of 350 to 550 °C, as brittleness could result. When these types of materials are cooled from tempering temperatures above 550 °C, the transition through this critical range should also be effected as rapidly as possible (see DIN 17022 Parts 1 and 2, [9] for additional information). This tempering sensitivity can be avoided by adding molybdenum or tungsten by alloying.

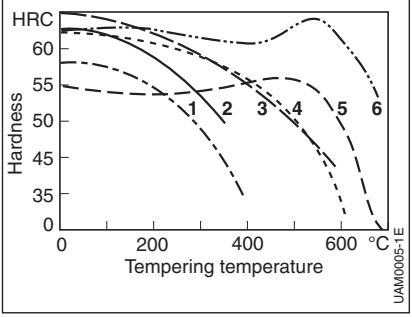
Quenching and drawing

Quenching and drawing involves a combination of hardening and tempering at a temperature which is generally between 540 °C and 680 °C. This procedure is designed to achieve an optimal relationship between strength and ductility. It is applied in cases where extreme ductility or malleability is required.

Particular care must be devoted to avoiding brittleness in the quench and draw operation (see above).

Figure 3: Tempering response of various types of steel

1 Unalloyed tempering steel (C45),
2 Unalloyed cold-working tool steel (C80W2),
3 Low-alloy cold-working tool steel (105WCr6),
4 Alloyed cold-working tool steel (X165CrV12),
5 Hot-working tool steel (X40CrMoV51),
6 High-speed tool steel (S6-5-2).



Annealing

Annealing can be applied to optimize certain operational or processing characteristics of parts. With this method, the parts are heated to a specific temperature and maintained there for an adequate period before being cooled to room temperature. The most important annealing processes in terms of technology are described in the following.

Stress-free annealing

Depending on the precise composition of the parts, stress-free annealing is carried out at temperatures ranging from 450 °C to 650 °C. The specific object is to achieve a reduction in internal stress in components, tools and castings.

Recrystallization annealing

Recrystallization annealing is applied with cold-form parts. The goal is to restructure the grain pattern in order to prevent increased hardening, thereby facilitating subsequent machining work.

The temperature requirement depends upon the composition of the material and the degree of deformation: it lies between 550 °C and 730 °C for steel.

Soft annealing and spheroidization

The purpose of soft annealing is to improve the machinability of material states which are difficult to cut or difficult to cold-form. The process involves heating the material in question to temperatures in excess of 600 °C, as briefly as possible above the Ac_1 temperature for the relevant steel, holding at this temperature, and slowly cooling to room temperature.

The temperature requirement is determined by the material's composition. It ranges from 650 °C to 850 °C for steel, and lower for nonferrous materials.

Spheroidization of cementite is applied when a microstructure with a granular

carbide pattern is desired. The cementite loses strength as a result of the annealing temperature and can pursue its striving for a body with as small a surface as possible (of the ball). If the initial structure is martensite or bainite, the result will be an especially homogeneous carbide distribution.

Normalizing
Normalizing is carried out by heating the parts to austenitizing temperature and then allowing them to gradually cool to room temperature. In low-alloy and unalloyed steels, the result is a structure consisting of ferrite and perlite. This process is essentially employed to reduce grain size, reduce the formation of coarse grain patterns in parts with limited reshaping, and to provide maximum homogeneity in the distribution of ferrite and perlite.

Precipitation hardening
This process combines solution treatment with aging at ambient temperature. The parts are heated and then maintained at a temperature to bring precipitated structural constituents into a solid solution, and quenched at room temperature to form a supersaturated solution. The aging process comprises one or several cycles in which the material is heated and held at above-ambient temperatures ("hot aging"). In this process, one or several phases, i.e. metallic bonds between certain base alloys, are formed and precipitated in the matrix.

The precipitated particles enhance the hardness and strength of the base microstructure. The actual characteristics are determined by the temperature and duration of the aging process (option of mutual substitution). Exceeding a certain maximum will usually reduce the strength and hardness of the final product.

Precipitation hardening is mostly applied for nonferrous alloys, but some hardenable steels (maraging steels) can also be processed.

Application
Steels which can be precipitation-hardened are used for example in rail-pressure sensors in common-rail systems.

Thermochemical treatment

In thermochemical treatment, the exchange of substances with suitable agents effects a change in the chemical composition of the base material. Specific function properties can be adapted by the diffusion of specific elements into the surface layer. Of particular importance for this process are the elements carbon, nitrogen and boron.

Carburizing, carbonitriding and case hardening

Carburizing increases the carbon content in the surface layer, while carbonitriding supplements the carbon enrichment with nitrogen. This process usually takes place at temperatures ranging from 750 to 1,050 °C in gases which give off carbon or nitrogen as a result of their disintegration by heat or excited in the plasma. The actual hardening is performed subsequently, either by quenching directly from the carburizing or carbonitriding temperature (direct hardening), or by allowing the parts to cool to room temperature (single hardening), or by allowing them to cool to a suitable intermediate temperature (e.g. 620 °C) prior to reheating (hardening after isothermic conversion). This process produces a martensitic surface layer, while the degree of martensite at the core is a function of hardening temperature, hardenability and part thickness.

Specific temperatures can be selected for either surface hardening in the upper layers with higher carbon content (case refining), or for the non-carburized core (core refining) (see DIN 17022 Part 3, [9]). Carburizing and carbonitriding produce a characteristic carbon declivity, with levels dropping as the distance from the surface increases (carbon curve). The distance between the surface and the point at which the mass carbon content is still 0.35 % is normally defined as the carburization depth.

The length of the carburizing or carbonitriding process depends upon the required carburization depth, the temperature and the atmosphere's carbon-diffusion properties.

Typical surface carbon contents during case hardening range between 0.5 and 0.85 % carbon by mass in order to achieve sufficient surface hardness. The carbon concentration essentially determines the surface hardness.

Excessive concentrations of carbon can lead to residual austenite or carbide diffusion, which could have negative effects on the performance of case-hardened parts in actual use. Control of the atmosphere's carbon level is thus extremely important in process management.

The most common carburizing processes used today are gas carburizing and vacuum carburizing. In the case of gas carburizing, the carbon level of the furnace atmosphere is regulated in such a way that the surface layer of the workpiece assumes the desired carbon content. A state of equilibrium is established with the surrounding furnace atmosphere. In the case of vacuum carburizing, on the other hand, the carbon content cannot be regulated. Here, the defined surface carbon content is adjusted by means of multistage carburizing. In a first stage, carburizing is performed to a very high surface carbon content in the range of material saturation. In the subsequent stage, this high carbon content is reduced by diffusion to the desired level. In practice, vacuum-carburizing processes consist of several consecutive carburizing and diffusion stages.

The gradient defining the relationship between hardness and depth corresponds to the carbon concentration curve. The case-hardening depth CHD can be taken from this. DIN EN ISO 2639 [18] defines this as the vertical distance to the surface to the layer which has a Vickers hardness of 550 HV 1.

The case-hardened part generally exhibits compression tension at the surface, and tensile stresses at the core. As with surface-hardened materials, this distribution pattern provides enhanced resistance to vibration loads.

In carbonitriding, nitrogen is also absorbed; it serves to improve the material's tempering properties, increase its durability and enhance its wear resistance. The positive effects are especially pronounced with unalloyed steels.

For additional, more detailed information on case hardening procedures, consult DIN 17022 Part 3 [9], and Information

Sheet 452 of the Steel Information Center, Düsseldorf [19].

Application
Injection nozzles subject to wear for Bosch common-rail systems which must be able to withstand high wear and internal compressive stresses are case-hardened by vacuum carburizing.

Nitriding and nitrocarburizing
Nitriding is a thermal treatment process (temperature range: 400 to 600 °C) which can be used to enrich the surface layer of virtually any ferrous material with nitrogen. In nitrocarburizing, a certain amount of carbon is diffused into the material to form nitrogen at the same time.

Molecular nitrogen, as is present in this temperature range in gaseous nitrogen, cannot diffuse into metallic materials. It is therefore necessary to offer diffusible nitrogen via suitable donor media. In practical applications, nitriding and nitrocarburizing processes are performed in gas atmospheres containing ammonia, in plasma containing nitrogen, or even in molten saline solutions containing cyanate. While ammonia gas releases diffusible nitrogen during its thermal disintegration, nitrogen is ionized in the plasma in order to split the molecules and facilitate the diffusion of nitrogen atoms.

The enriching nitrogen in the surface layer causes the precipitation of nitrides, in response to which the surface layer hardens. In the final analysis, this results in greater resistance to wear and corrosion, and in greater endurance strength.

As the process employs relatively low temperatures, there are no volumetric changes of the kind associated with transformations in the microstructure, so that changes in dimensions and shape are minute.

The nitrided region consists of an outer layer, several millimeters in depth, and a transitional white layer, the hardness of which may be anywhere from 700 HV to over 1,200 HV, depending on the composition of the material. Still deeper is a softer diffusion coating extending several tenths of a millimeter in which the nitrogen content decreases as the distance from the surface increases. The thickness of the individual layers is determined by the temperature and duration of the treatment process. The process produces a hardness gradient (similar to that which results from surface and case hardening); this gradient furnishes the basis for determining the nitriding hardness depth NHD. DIN 50190 Part 3 [20] defines this as the vertical distance from the surface to the point where the hardness corresponds to a defined limit. This limit hardness is usually actual core hardness + 50 HV 0.5.

The material's resistance to wear is essentially determined by the white layer, which contains up to 10 mass components of nitrogen in %. The nitriding hardness depth and the surface hardness determine the material's resistance to alternating cyclic stress (for additional details, see DIN 17022 Part 4, [9], and Information Sheet 447 from the Steel Information Center, Düsseldorf, [21]).

The corrosion resistance of nitrided or nitrocarburized workpieces can be significantly increased by postoxidation in water vapor or other suitable gases, or in molten saline solutions at temperatures ranging between 350 °C and 550 °C.

Application
Injection nozzles which are used, for example, in Bosch common-rail systems and which must be able to withstand extreme operating temperatures are gas-nitrided.

Components for windshield and rear-window wiper systems are nitrocarburized and postoxidized to increase their resistance to corrosion and wear. This is how these components get their typical black color.

Boron treatment
Boron treatment involves enriching the surface layer of ferrous materials with boron. Depending upon duration and temperature (normally 850 to 1,000 °C) of the treatment, a white layer of 30 µm to 0.2 mm in depth with a hardness of 2,000 to 2,500 HV is produced. It consists of iron-boron.

Boron treatment is particularly effective as a means of protecting against abrasive wear. However, the comparatively high process temperature leads to relatively

large changes in shape and dimensions, meaning that this treatment is only suitable for applications in which large tolerances can be accepted.

Application
Partly boron-treated tool holders with high wear resistance are used in Bosch hammer drills.

References
[1] DIN EN ISO 18265: Metallic materials – Conversion of hardness values.
[2] DIN EN ISO 6508: Metallic materials – Rockwell hardness test.
[3] DIN EN ISO 6506: Metallic materials – Brinell hardness test.
[4] DIN EN ISO 4498: Sintered metal materials, excluding hardmetals – Determination of apparent hardness and microhardness.
[5] DIN EN ISO 6507: Metallic materials – Vickers hardness test.
[6] DIN EN ISO 4545: Metallic materials – Knoop hardness test.
[7] DIN EN ISO 14577: Metallic materials – Instrumented indentation test for hardness and materials parameters.
[8] DIN EN 10052: Vocabulary of heat treatment terms for ferrous products.
[9] DIN 17022: Heat treatment of ferrous materials – Methods of heat treatment.
Part 1: Hardening, austempering, annealing, quenching, tempering of components.
Part 2: Hardening and tempering of tools.
Part 3: Case hardening.
Part 4: Nitriding and nitrocarburizing.
Part 5: Surface hardening.
[10] DIN EN 10083: Steels for quenching and tempering. Part 1: General technical delivery conditions.
Part 2: Technical delivery conditions for non-alloy steels.
Part 3: Technical delivery conditions for alloy steels.
[11] DIN EN 10085: Nitriding steels – Technical delivery conditions.
[12] DIN EN ISO 4957: Tool steels.
[13] DIN EN ISO 642: Steel – Hardenability test by end quenching (Jominy test).
[14] DIN 17021: Heat treatment of ferrous materials.
[15] DIN EN ISO 18265: Metallic materials – Conversion of hardness values.
[16] DIN EN 10328: Iron and steel – Determination of the conventional depth of hardening after surface hardening.
[17] DIN 17021-1: Heat treatment of ferrous materials; material selection, steel selection according to hardenability.
[18] DIN EN ISO 2639: Steels – Determination and verification of the depth of carburized and hardened cases.
[19] Information Sheet 452 of the Steel Information Center, Düsseldorf: “Einsatzhärten”, Edition 2008.
[20] DIN 50190-3: Hardness depth of heat-treated parts; determination of the effective depth after nitriding.
[21] Information Sheet 447 of the Steel Information Center, Düsseldorf: “Wärmebehandlung von Stahl – Nitrieren und Nitrocarburieren”, Edition 2005.

Corrosion and corrosion protection

Corrosion processes

Corrosion is the damage to a metal as a result of a reaction with substances in the environment. Corrosion processes always contain interphase reactions. An example of this type of reaction is metal scaling, i.e. oxidation in hot gases. As it proceeds, metal atoms oxidize to form nonmetallic compounds in a process emanating from the affected material's surface. In thermodynamic terms, the process can be viewed as an entropic transition from an ordered, high-energy state into a less-ordered state of lower energy, and consequently greater stability. The process takes place of its own accord.

The following deals exclusively with the corrosion that occurs at the phase boundary between the metal and aqueous phases (electrolyte), generally referred to as electrochemical corrosion.

Corrosive attack

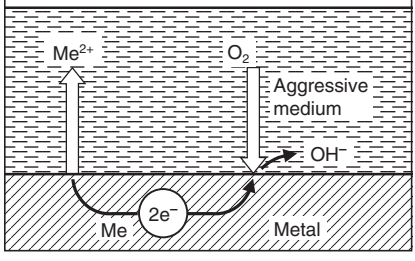
Anodic subprocess

Two essentially distinct reactions occur in electrochemical corrosion: in the anodic subprocess, the directly visible corrosion process, the metal passes into the oxidized state as described in the reaction equation

$$Me \rightarrow Me^{n+} + n\ e^-$$

freeing an equivalent number of electrons (Figure 1). The metal ions thus formed can either be dissolved in the electrolyte, or can precipitate out on the metal as corrosion products (e.g. rust) after reacting with constituents in the attacking medium [1].

Figure 1: Free corrosion at the metal/corrosive-medium phase boundary.
In an aggressive medium (neutral or alkaline) oxygen is reduced at the corroding metal. Corrosion products are formed simultaneously.



Cathodic counter-reactions

This anodic subprocess can continue only as long as the electrons it produces are consumed in a second process. This second subprocess is a cathodic subreaction. In neutral or alkaline media oxygen is reduced in accordance with

$$O_2 + 2\ H_2O + 4\ e^- \rightarrow 4\ OH^-$$

to hydroxyl ions, which in turn are able to react with the metal ions. In acidic media the hydrogen ions are reduced via the formation of free hydrogen, which escapes as a gas according to the following formula:

$$2\ H^+ + 2\ e^- \rightarrow H_2\ .$$

Electrochemical series of metals

Metals are often ranked in "electrochemical series of metals" corresponding to consecutively higher standard potentials (see also Electrochemistry). As the value of standard potential increases the metals are referred to as "more noble", and as the value decreases as "more base".

It should be emphasized that the table ("Standard electric potentials of metals") is limited to thermodynamic values, and does not reflect the effects of corrosion kinetics, for example as encountered in the formation of protective layers. Most corrosion-resistant structural materials, however, develop protective layers, which are crucial to corrosion resistance. Thus, for example the materials aluminum, zinc and titanium seem to be very base metals, but are highly resistant due to the formation of protective layers. The formation of protective layers is exploited in the field of electrochemical corrosion protection.

Types of corrosion

General surface corrosion
General surface corrosion is a type of corrosion with uniform attrition of material over the entire boundary surface between the material and the attacking medium. This is a very frequent type of corrosion in which the material penetration rate (removal depth) can be calculated per unit of time based on the corrosion current.

Pitting corrosion
Pitting corrosion is a limited localized attack by a corrosive medium which penetrates the material by forming holes, or pits, whose depth is almost always greater than their diameter. Practically no material is removed from the surface outside the pitted areas. Pitting corrosion is frequently caused by chloride ions (e.g. from table salt).

Contact corrosion
When two different metals moistened by the same medium are in mutual electrical contact, a cathodic subprocess occurs at the more noble metal, while the anodic subprocess progresses at the baser material. This is called contact corrosion.

Crevice corrosion
Crevice corrosion is a corrosive attack primarily occurring in narrow crevices, caused by concentration differences in the corrosive medium, e.g. as a result of long oxygen diffusion paths. This type of corrosion generates potential differences between the crevice extremities, leading to intensified corrosion in more poorly ventilated areas.

Stress corrosion cracking
This is corrosion stemming from the simultaneous concerted action of a corrosive medium and mechanical tensile stress (which can also be present as internal stress in the object itself). Intergranular or transgranular fissures form, in many cases without the appearance of visible corrosion products.

Vibration corrosion cracking
Vibration corrosion cracking is corrosion caused by the simultaneous effects of a corrosive medium and mechanical fatigue stress, e.g. caused by vibrations. Transgranular cracks are formed, frequently without visible deformations.

Intergranular and transgranular corrosion
This involves types of corrosion characterized by selective formation along the grain boundaries or roughly parallel to the deformation plane in the grain interior.

Dezincification
Selective dissolution of zinc from brass, leaving behind a porous copper structure.

Corrosion testing

Electrochemical testing procedures
The primary tool for electrochemical corrosion testing is the potentiostat. Here metals or other electrically conductive materials are analyzed in different corrosive media. The typical setup with three electrodes is shown in Figure 2. This arrangement is essentially used to determine – as well as the corrosion potentials of the corroding materials – primarily the corrosion currents. In the case of uniform surface corrosion, the corrosion-current parameters are then used for the definition of the attrition mass and removal depth per unit of time.

The electrochemical testing procedures are a valuable supplement to the non-electrochemical methods, because they provide an understanding of corrosion mechanisms that occur. In addition to the small amount of corrosive medium required, another advantage of electrochemical procedures over non-electrochemical methods is that they provide quantitative data on attrition rates.

The potentiostat is used to alter the potential of the material to be analyzed and measure the flowing current. Evaluation enables reaction mechanisms and above all corrosion current densities to be determined.

Polarization-resistance measurement
In the case of free corrosion the corrosion current is determined from the polarization resistance (slope of the cumulative current-density/potential curve); testing entails subjecting the metal to minimal, alternating anodic and cathodic pulses.

Impedance spectroscopy
Electrochemical Impedance Spectroscopy (EIS) is employed to examine corrosion mechanisms. This alternating-current technique determines the AC resistance (impedance) of an electrochemical test object as a function of frequency. A low-amplitude sinusoidal alternating voltage is superimposed on the working electrode's potential, and the current response is measured. To interpret the measurement the system is approximated in the form of an equivalent network. By way of example, Figure 3 shows the equivalent network for the system, consisting of metal, coating, and medium.

Impedance elements (resistances, capacitances, inductances) are assigned to the sequences (e.g. electron transfer) at the phase boundary. In this simple example from Figure 3 the coating is – along similar lines to the phase boundary of material to corrosive medium – described as capacitor and resistor.

Figure 2: Schematic setup for measuring the current density/potential curves with a potentiostat
1 Working electrode,
2 Reference electrode,
3 Counter-electrode.

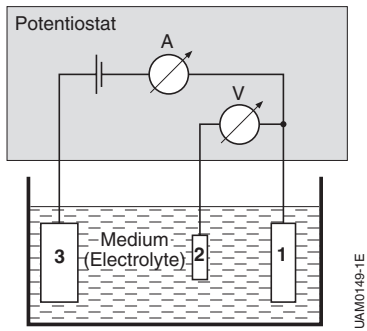
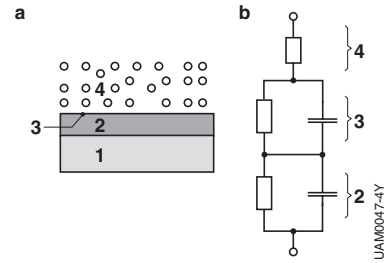


Figure 3: Evaluation of EIS data (electrochemical impedance spectroscopy)
a) Arrangement,
b) Equivalent network.
1 Metal,
2 Coating,
3 Phase boundary metal to medium,
4 Corrosive medium.



Direct conclusions can then be drawn from the equivalent networks and the impedance elements about various characteristics, such as the effectiveness of corrosion-protection measures, porosity, thickness, a coating's water-absorption ability, the effectiveness of inhibitors, the corrosion rate of the base metal, and so on.

Contact-corrosion current measurement
When contact corrosion is measured the two affected metals are immersed in the same corrosive medium. During measurement they are not directly contacted, but instead connected via the potentiostat. Here the contact-corrosion potential and the contact-corrosion current flowing between the two metals are directly measured and the curve plotted against the measurement period.

Non-electrochemical corrosion-testing procedures
In non-electrochemical testing procedures test specimens essentially are first exposed to a corrosive environment (exposure tests) and then the change in the test specimens is assessed.

Simple standard tests (e.g. corrosion load by neutral salt spray fog testing) are conducted with the aim of testing quality or comparing corrosion-protection measures and materials. The emphasis is placed on using tests that can uncover weak points in corrosion-protection measures and describe different qualities.

After corrosion load is completed, it is possible by comparing images in accordance with DIN EN ISO 4628-3 [2] to assign rust levels (degrees of rusting) which are defined according to rust coverage or surface perforation (Table 1).

Table 1: Rust level and rust surface acc. to DIN EN ISO 4628-3

Rust level	Rust surface in %
R_i0	0
R_i1	0.05
R_i2	0.5
R_i3	1
R_i4	8
R_i5	40...50

For the purpose of service-life assessment, testing procedures have evolved to reflect specific requirements, e.g. motor-vehicle testing. These are as a rule corrosion loads with cyclically alternating loading by salt spray fog, dry phase and moist phase. These tests provide reliable indices of projected service life under normal operating conditions by using short-term exposure in extremely harsh conditions to simulate long-term stresses in the real world (e.g. testing in accordance with Bosch standard N42AP 226 “Climatic tests – Tightened life – Corrosion testing”).

For a further adaptation of corrosion load to practical operation products are operated under tightened conditions (e.g. vehicle tests of the different automobile manufacturers).

In addition to the corrosion loads for life testing the focus is on assessing the function of the products.

Corrosion protection

The manifestations and mechanisms of corrosion are many and varied, so widely differing methods can be adopted to protect metals against corrosion attack. Corrosion protection means intervening in the corrosive process with the object of reducing the rate of corrosion in order to prolong the service life of the components.

Corrosion protection can be achieved by applying four basic principles:

– Measures in planning and design: the choice of suitable materials and the suitable structural design of components.
– Measures that intervene in the corrosive process by electrochemical means.
– Measures that separate the metal from the corrosive medium by protective layers or coatings.
– Measures that influence the corrosive medium, for example, the addition of inhibitors to the medium.

Corrosion protection by means of suitable structural design

Material selection

Selecting suitable materials which feature optimum resistance to corrosion under the expected conditions can be of considerable assistance in avoiding corrosion damage. When the costs that would otherwise be incurred for upkeep and repair are factored into the long-term cost of ownership equation, a more expensive material can often be the more cost-effective alternative.

Design

Design measures, too, are of major importance. A great deal of skill and expertise goes into design, particularly regarding the connections between parts that are made of the same material or different materials.

Corners and edges of sections are difficult to protect, and this is where corrosion can easily attack. A favorable installation position can avoid corrosion (Figure 4).

Beads and welts can trap dirt and moisture. Suitable surfaces and drain openings can help avoid this problem (Figure 5).

In the case of welds the structure is altered in a detrimental way. In order to avoid crevice corrosion, welds have to be smooth and free of gaps (Figure 6).

Contact corrosion can be avoided by joining same or similar metals, or by installing washers, spacers, or sleeves to ensure that both metals are electrically insulated (Figure 7).

Figure 4: Good and bad installation positions for profile sections

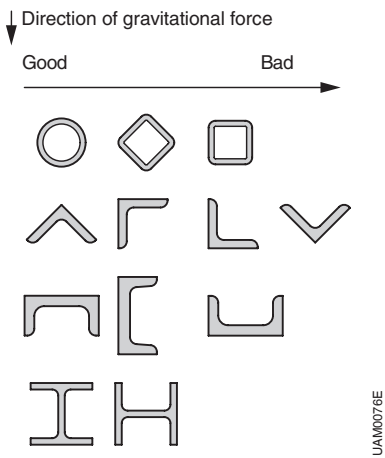
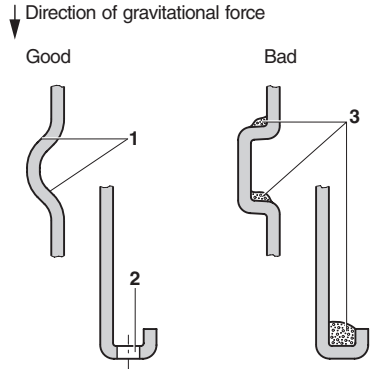


Figure 5: Design of beads and welts

1 Sloping faces (foreign matters slide off),
2 Wide gap with drain,
3 Deposits of dirt and moisture.



Electrochemical processes
The schematic current-density/potential curves for a metal suitable for passivation coating (e.g. by oxide layer on stainless steel) show how these processes work (Figure 8). The current-density values arranged in ascending order on the y-axis represent anodic currents corresponding to the corrosion reaction defined in the equation

$Me \rightarrow Me^{n+} + n\ e^{-}$.

In contrast, the descending current-density values represent cathodic currents.

As can be seen from the diagram, corrosion only occurs when potentials $E_a \leq E \leq E_p$ and $E \geq E_d$. Therefore corrosion can be suppressed by applying an external voltage. There are two basic ways of doing this.

Cathodic protection
For cathodic protection, the potential is shifted so far toward the left that no anodic currents flow, leading to $E < E_a$. As an alternative to applying voltage from an external source, it is also possible to shift the potential by contacting a base metal to act as a "sacrificial" reactive anode.

Anodic protection
Another option is, by applying an external voltage, to shift the potential of the threatened electrode into the passive range, i.e. into the potential range between E_p and

Figure 6: Design-related crevices in welds and how they can be avoided

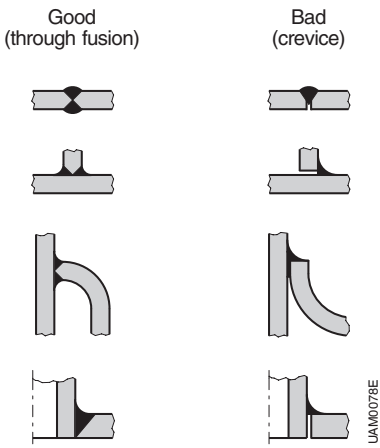


Figure 7: Electric insulation to avoid contact corrosion
a) Screwed/bolted joint,
b) Retainer,
c) Riveted joint.
1 Insulating washer,
2 Insulating spacer.

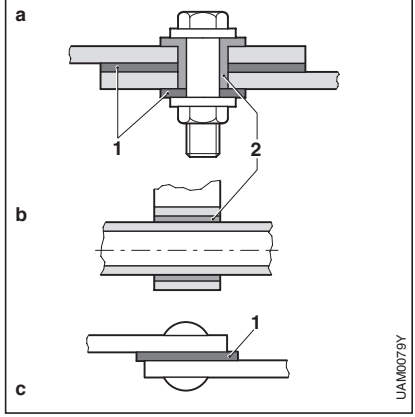
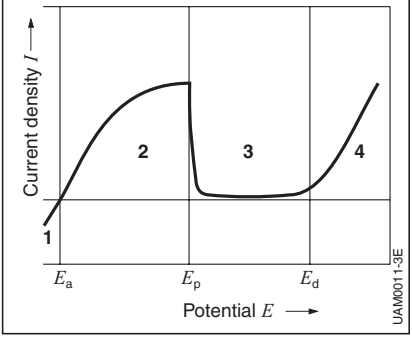


Figure 8: Schematic curve of current density as a function of the potential for a metal which can be passivated
1 Cathodic protection, occurs when $E < E_a$,
2 Corrosion without protection,
3 Anodic protection, occurs when $E_p < E < E_d$,
4 Transpassive range.
E_a Free corrosion potential of the metal in an active state,
E_p Passivating potential,
E_d Breakthrough potential.



E_d. This is called anodic protection. The anodic currents, which flow in the passive range, are less than those in the active range by exponential powers of between three and six, depending on the type of metal and the corrosive medium. The result is excellent protection for the metal.

However, the potential should not exceed E_d, as oxygen would be produced in this transpassive range, potentially leading to higher rates of oxidation. Both of these effects would cause the current to increase.

Coatings

Coatings inhibit corrosion by forming protective films which are applied to the surface of the metal to be protected, where they resist attack by the corrosive medium (see also Deposits and coatings).

Inhibitors

Inhibitors are substances added to the corrosive medium in low concentrations (up to a maximum of several hundred ppm) for absorption at the surface of the protected metal. They drastically reduce the rate of corrosion by blocking either the anodic or the cathodic subprocess (frequently blocking both subprocesses simultaneously). Organic amines and amides of organic acids are the most frequent inhibitors. In automotive applications, for example, inhibitors are used in fuel additives. They are also added to antifreeze in order to inhibit corrosion damage in the coolant circuit.

Vapor-phase inhibitors

Vapor-phase inhibitors (VPI), or volatile corrosion inhibitors (VCI), are organic substances of moderate vapor pressure. They are frequently enclosed in special-purpose packing materials or as solvents in liquids or emulsions with oil. The inhibitors evaporate or sublimate over the course of time, and are adsorbed as monomolecules on the metal, where they inhibit (i.e. slow down or stop) either anodic or cathodic corrosion subreactions, or both at once. Dicyclohexylamin nitrite is a typical example.

For optimal effectiveness, the inhibitor should form a sealed coating extending over the largest possible surface area. Standard commercial vapor-phase inhibitors are generally a combination of numerous components capable of providing simultaneous protection for several metals or alloys; exceptions: cadmium, lead, tungsten and magnesium.

Vapor-phase inhibitors provide only temporary protection for metallic products during storage and shipping. They must be easy to apply and remove. The drawback of vapor-phase inhibitors is that they are potentially injurious to health.

References

[1] DIN 50919:2016-02: Corrosion of metals – Corrosion investigations of bimetallic corrosion in electrolytic solutions.

[2] DIN EN ISO 4628-3:2016-07: Paints and varnishes – Evaluation of degradation of coatings – Designation of quantity and size of defects, and of intensity of uniform changes in appearance.
Part 3: Assessment of degree of rusting (ISO 4628-3:2016); German version EN ISO 4628-3:2016.

Deposits and coatings

Deposits and coatings are used to adapt the surface properties of components to particular requirements ("surface engineering", Table 1). It is thus possible, for example, for a component to be made from a tough, low-cost material which still has a hard and wear-resistant surface. The main applications for deposits and coatings are

– Corrosion protection (as well as functional retention, often for decorative reasons)
– Wear protection
– Joining and bonding techniques (plug-in, welded, soldered, bonded, and crimped contacts).

In coating systems, distinctions are made between the following types of coating:

– Deposits in which one layer is applied
– Conversion coatings in which the functional coating is produced by chemical or electrochemical conversion of the base material, and
– Diffusion coatings in which the functional coating is produced by diffusion of atoms or ions into the base material.

Deposits

Electrodeposits

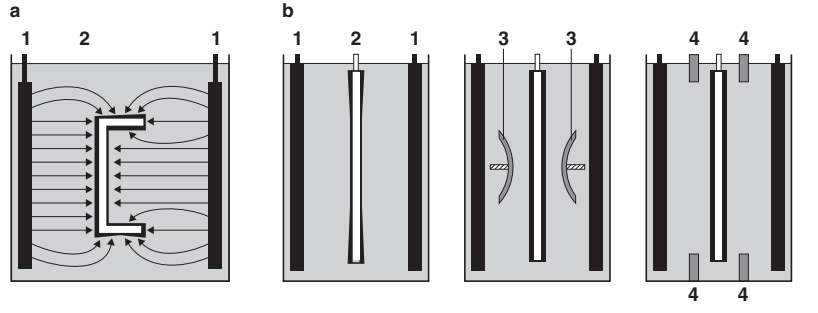
Electrodeposits are deposited using an external power source. The workpieces to be coated are immersed into an electrolyte (Figure 1) which contains ions of the metal to be deposited. The metal ions consumed during the deposition process are either supplied by the anode solution or added to the electrolyte (when inert anodes are used). The progression and distribution of the field lines influence the liner-thickness distribution. A uniform layer thickness can be achieved by means of an optimized configuration of anodes and screens.

Electrodeposits are widely used in corrosion protection, in wear protection and for electrical contacts. Some important deposits are described below.

Zinc and zinc alloys

Electrolytically deposited zinc coatings are widely used as corrosion protection for steel components and are a cathodically effective means of corrosion protection (sacrificial effect). To increase the corrosion-protection effect, zinc deposits are passivated in a chrome(VI)-free solution. Zinc alloys, such as zinc-nickel with

Figure 1: Electrodeposits
a) Progression and distribution of electrical field lines between anode and cathode as caused by nonuniform liner-thickness distribution.
b) Comparison of liner-thickness distribution during electroplating without auxiliary agent, with auxiliary anodes and with screens (schematic representation).
1 Anode, 2 Cathode, 3 Auxiliary anode, 4 Screen.



approx. 15 % nickel offer a significantly higher degree of corrosion protection.

Nickel
Electrolytically deposited nickel coatings offer limited corrosion protection with an attractive visual appearance. One area of automotive engineering in which this is chiefly used is the nickel-plating of spark-plug shells.

Chrome
In the case of chrome deposits, a distinction is made between hard chrome and bright chrome. Bright chrome is used as a deposit approx. 0.3 µm thick with a nickel or copper/nickel intermediate deposit. In the past, bumpers and molding strips were provided with bright-chrome deposits for decorative reasons. Decorative chrome deposits have become an important element in vehicle design.

Hard-chrome deposits are chrome deposits with a thickness greater than 2 µm. Owing to the high level of hardness of the electrolytically deposited chrome coating, it is ideally suited for use as a wear-protection coating. In the past, hard-chrome deposits were often applied in strong thicknesses and then mechanically reworked. Due to further developments in process engineering, components today are increasingly custom-coated with coating thicknesses ranging from 5 to 10 µm, and then used without reworking (e.g. components for fuel injectors).

Tin
Electrolytically deposited tin coatings are mainly used as contact surfaces for plug-in and switch contacts, and as solder contact surfaces. A coating thickness of 2 to 3 µm is ideal for plug-in contacts. Greater coating thicknesses are required for solder applications in order to ensure solderability even after extended storage periods.

Gold
Gold deposits are normally used for contacts subject to stringent requirements. They are characterized by good conductivity, low contact resistance, and good resistance to corrosion and pollutant gases. This ensures contact stability. Hard-gold deposits (with approx. 0.5 % alloying constituents, predominantly cobalt) are harder and more abrasion-resistant than pure-gold deposits and are suitable for contacts subject to mechanical load.

Chemically deposited coatings (no external current)
In contrast to electrodeposits, chemically deposited coatings are characterized by a more uniform coating-thickness distribution on the component, since deposition in a coating bath does not occur under

Table 1: Areas of application for coatings

Coating system	Coating material	Main application
Electrodeposits	Zn, ZnNi, Ni Cr Sn, Ag, Au	Corrosion protection Wear protection Decorative Electrical contacts
Chemical deposits	NiP, NiP dispersions Cu, Sn, Pd, Au	Wear and corrosion protection Electronic applications
Hot-immersion deposits	Zn, Al Sn	Corrosion protection Electrical contacts
Paints (wet and powder-based paints)	Organic polymers Pigments (color particles)	(Decorative) corrosion protection Wear reduction Electrical insulation
PVD and CVD coatings (plasma technology)	TiN, TiCN, TiAlN DLC i-C(WC), a-C:H	Wear protection of tools Reduction of friction and wear of components

the influence of an external electric field. Owing to the slow deposition rate and the costly chemicals in the coating bath, they are more expensive than electroplated liners. The following have become widespread in industry:

– "Chemical nickel" (nickel-phosphorus, NiP) as a corrosion- and wear-protection deposit, used for example in electric fuel pumps or in high-pressure pumps for gasoline direct injection.
– "Chemical copper" and "Chemical tin" in printed-board technology.

Hot-immersion deposits

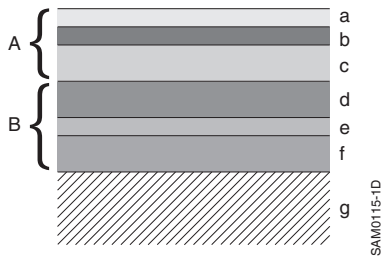
Hot-immersion coatings are deposited by immersing the substrates in a molten metal bath. Cathodically effective hot-immersion deposits are used to protect low-alloyed steels against corrosion. Zinc, zinc/aluminum alloys and aluminum are used. Coating start material such as sheet and strip is a cheap and widely used practice. However, free punch edges must be accepted here.

Tin and tin-alloy coatings deposited in hot-immersion priming are primarily used as surfaces for plug connectors and as soldering surfaces.

Figure 2: Layered structure of paint coatings on car bodies

A Visual appearance,
B Corrosion protection.
a Top coat, transparent,
b Top coat, colored,
c Extender,
d Cathodic deposition (CD),
e Phosphate coat,
f Zinc electrodeposit,
g Sheet steel.

Paint structure (coats, see Table 2):
Solid: 1, 2, 3a, 3b, 4, 5.
Metallic: 1, 2, 3c, 3d, 4, 5.



Zinc-lamella deposits

Zinc-lamella deposits are coatings based on zinc and aluminum lamellas in an inorganic binder. They are applied by immersion centrifuging or electrostatic spraying and thermal hardening. Zinc-lamella deposits are low-cost corrosion-protection coatings for mass-produced steel parts (e.g. screws/bolts).

Paints

Paints have a wide variation range because of their broad chemical bases and numerous means of application, such as brushing, spraying, and immersing, even with the use of current (CD: cathodic deposition).

Paint coatings can also fulfill a multitude of different functions. In motor vehicles, the primary function fulfilled by paint coatings is corrosion protection accompanied by a decorative effect, but other functions include wear protection, soundproofing or electrical insulation.

The very high demands imposed on corrosion protection and visual appearance over the service life of a vehicle are guaranteed by a complex, elaborate layered structure (Figure 2 and Table 2). The assemblies in the engine compartment, on the other hand, are usually only given one or two coats of paint. Less importance is placed here on decorative properties.

Low-solvent systems, above all water-based paints, are almost always used in automotive engineering. Where possible, powder-based paints and UV-hardening paints which are completely solvent-free are used.

Table 2: Structure of solid-color and metallic coatings

Layer	Layer thickness	Structure	Composition					
	in µm		Binders	Solvents	Pigments	Extenders	Additives and SC	Applications
1	20 to 25	Cathodic deposition	Epoxy resins Polyurethane	Water, small amounts of water-miscible organic solvents	Inorganic (organic)	Inorganic extenders	Surface-active substances, anti-crater agents, SC 20 %	Electrophoretic coating
2a	approx. 35 to 50	Extender	Polyester, melamine, urea, and epoxy resins	Aromatic compounds, alcohols	Inorganic and organic	Inorganic solids	E.g. wetting agents, surface-active substances SC 55…>70 %	PS ESTA-HR
2b	approx. 35 to 50	Water extender	Water-thinnable polyester, polyurethane, and melamine resins	Water, small amounts of water-miscible organic solvents			SC 43 to 50 %	PS ESTA-HR
3a	40 to 50	Solid-color top coat	Alkyd and melamine resins	Esters, aromatic compounds, alcohols		–	E.g. leveling, wetting agents SC 50 to 60 %	PS ESTA-HR
3b	10 to 35 (color-dependent)	Water-thinnable solid-color base coat	Water-thinnable polyester, polyurethane, polyacrylate, melamine resins	Small amounts of water-miscible cosolvents		–	Wetting agents SC 20 to 40 %	PS ESTA-HR
3c	10 to 15	Metallic, base coat	CAB, polyester, melamine resins	Esters, aromatic compounds	Aluminum and mica particles	–	SC 15 to 30 %	PS ESTA-HR
3d	10 to 15	Water-thinnable metallic base coat	Water-thinnable polyester, polyurethane, and polyacrylate, and melamine resins	Small amounts of water-miscible cosolvents	Aluminum mica particles Organic and inorganic pigments	–	Wetting agents SC 18 to 25 %	PS ESTA-HR
4a	40 to 50	Clear coat conventional	Acrylic and melamine resins	Aromatic compounds, alcohols, esters	–	–	E.g. leveling agents and light stabilizers, SC 45 %	PS ESTA-HR
4b	40 to 50	2C-HS	HS acrylate resin polyisocyanates	Esters, aromatic compounds	–	–	E.g. leveling agents and light stabilizers, SC 58 %	
4c	40 to 50	Powder-slurry clear coat	Urethane-modified epoxy/carboxy system	–	–	–	E.g. light stabilizers, SC 38 %	
4d	40 to 50	Powder clear coat	Acrylate resin	–	–	–	100 %	ESTA PS

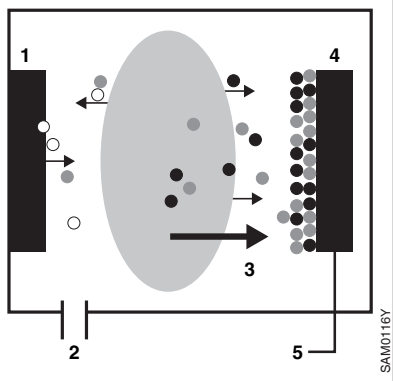
Acronyms: ESTA-HR Electrostatic high rotation; SC Solids content; PS Pneumatic spray; 2C-HS 2-components high solid (high levels of non-volatile matter)

PVD and CVD coatings
PVD and CVD coating systems are generally deposited in a vacuum supported by plasma or heat treatment on components and tools (Figure 3). There are two different procedures, depending on whether the coating-forming material is obtained from the solid (PVD, physical vapor deposition) or gaseous (CVD, chemical vapor deposition) phase. Most modern procedures combine the two categories.

Hard-material coatings
Tools are coated with hard material to increase their service life and their performance. A typical example of the tool coatings widely introduced onto the market is the gold-colored titanium nitride (TiN), which is produced, for example, by the cathodic sputtering or arc evaporation of titanium in a reaction with nitrogen. More recent coating systems, such as titanium carbonitride (TiCN) or titanium aluminum nitride (TiAlN), can be used for high-speed cutting or for dry machining.

Figure 3: Schematic of plasma-assisted PVD/CVD coating in vacuum chamber
1 Cathode with sputter target,
2 Gas inlet,
3 Intensive ion bombardment,
4 Substrate,
5 Substrate tension

○ Argon,
● Metal/carbon,
● Nitrogen/hydrogen.



Superhard materials, such as diamond, are increasingly used to coat carbide tools.

DLC coatings
Protecting components against wear is subject to special conditions. Here, the coating should establish a low coefficient of friction for the components that come into contact with and move relative to each other and minimize the wear for the entire layout. DLC (diamond-like carbon) coatings protect not only the coated component against wear, but also the uncoated opposed body. They have a very low coefficient of friction of 0.1 to 0.2, both in a dry friction pairing against steel and under media. DLC layers are media-resistant and have a corrosion-inhibiting effect. Due to these advantages, DLC coatings are very important in the field of coating components. However, it must be noted that coatings containing hydrogen will degrade in an oxidizing atmosphere at a local temperature of 350 °C and higher. Hard-material coatings, such as TiN, can withstand considerably higher temperatures and are also used in component coating.

Different coating compositions and processes allow materials to be adapted to different wear loads or combinations of abrasive, vibration, sliding, seizure, and adhesive wear. The term DLC therefore describes not a layer but a class of materials within which there exists a multitude of coating variants with all manner of properties and with manufacturer-specific nomenclature.

Plasma-assisted PVD and CVD processes are conducted in such a way that the tempering temperature of the component is not exceeded during coating. This avoids impairing the hardness of the steel base material.
Low-friction carbon coatings containing metal carbide i-C(WC) are electrically conductive and have a microhardness of approx. 1,800 HV at a modulus of elasticity of 150 to 200 GPa.

In comparison, metal-free carbon coatings (a-C:H) offer increased hardness of approx. 3,500 HV and significantly improved wear resistance, but this is ac-

companied by an increase in brittleness. They are electrically insulating.

The greatest coating hardness and wear resistance for DLC coatings is currently achieved with metal- and hydrogen-free ta-C (tetrahedral amorphous carbon). For all DLC coatings, in view of their high internal compression stress, it is important to ensure a stable connection to the substrate, which is achieved for the most part with a metallic adhesive coating system.

With thicknesses of 2 to 4 µm, diamond-like carbon coatings offer very good wear protection and are especially suitable for precision components subject to high mechanical loads. There is no rework needed after coating.

Thermal spray-coating
In thermal spraying metals or alloys are melted thermally or in a plasma flame and sprayed onto the surface to be coated. There are a multitude of ways in which the procedure is modified in which ceramic surfaces can be created as well as metallic coatings. By choosing the relevant procedure, it is possible to create tightly sealed coatings for example for wear or corrosion protection or porous, permeable coating systems.

Coatings

Diffusion coatings
Surface treatment can be selectively combined with surface hardening by using the diffusion process to thermochemically carburize, carbonitride, or chromate the metal, or treat it with boron or vanadium (see Heat treatment of metallic materials, Thermochemical treatment). The metal can also be oxidized, nitrided, or sulfided without hardening.

Conversion coatings
Conversion coatings are formed not exclusively by the application of a material, but by in part the chemical or electrochemical conversion of the base material.

Browning coatings
Browning coatings consist of thin iron-oxide layers (predominantly Fe_3O_4) which are formed by the oxidation of the iron in steel in an alkaline, aqueous solution containing nitrite at temperatures greater than 100 °C. With subsequent oiling, they offer temporary corrosion protection for low-alloy steels.

Phosphating coatings
Phosphating coatings are formed on steel, galvanized steel, and aluminum in solutions containing phosphoric acid by immersion or spraying. Zinc-phosphate coatings are predominantly used as a wash primer for paint coatings or in combination with anti-corrosion oil. Manganese phosphate serves as a wear-protection coating with anti-seizure properties, and as a wash primer for other coatings to improve component antifriction properties.

Anodized coatings
Anodized coatings are formed by the electrochemical conversion of metal into metal oxide in aqueous electrolytes. Aluminum, magnesium, and titanium can be anodized. Anodized coatings on aluminum materials are widely used for corrosion and wear protection.

Tolerances

Form deviations

No component can be manufactured with absolute precision, i.e. with an ideal geometry. A whole range of causes is responsible for this, e.g., deformations as a result of machining forces, the play in guides of machine tools, tool wear, vibrations, and much more. The consequence of this is that every component exhibits dimensional, geometrical and surface deviations. To guarantee the function it is necessary that all the permissible deviations be indicated in the drawing.

Dimensional deviations
Each dimension can only be manufactured between an upper and a lower limit dimension. The difference between these two limits is called the dimensional tolerance.

Geometrical deviations
Geometries also exhibit deviations from the ideal geometry. A distinction is made between form and positional deviations.

Surface deviations
The surface of a workpiece likewise cannot be manufactured to the ideal smoothness levels. The description of the surface texture can be specified with the aid of the primary, waviness and roughness profiles.

Geometrical Product Specifications

Geometrical Product Specifications (GPS) refers to all the indications in a CAD model or a drawing which are required to manufacture and check (verify) a product. Product specifications must be carried out in such a way that the component to be manufactured performs its function without limitations.

The aim of Geometrical Product Specifications is to provide a clear and complete description of the product in accordance with standardized rules. These rules are laid down in the form of GPS standards. In addition, the design intent, i.e. what is important for the function, must be apparent from a drawing.

Important standards which are important for the correct interpretation of tolerances are set out in ISO 8015 [1]. These include among others the standard tolerancing principle.

Tolerancing principles
The relationship between the dimensional and geometrical tolerances is defined by the tolerancing principle. According to GPS there are two possibilities for this.

1. The envelope principle
Until 1985 no tolerancing principle was indicated on drawings, but was instead interpreted according to the Taylor principle. This corresponds to the envelope requirement at which the maximum

Figure 1: Envelope requirement
1 Envelope with maximum material size.

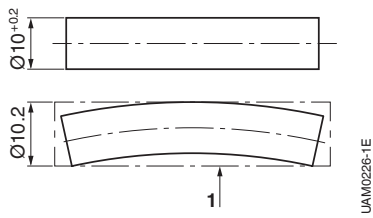
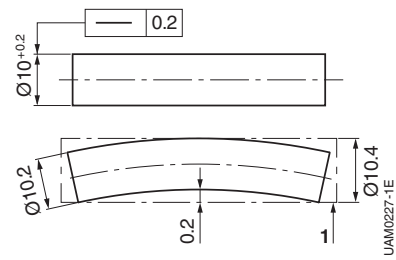


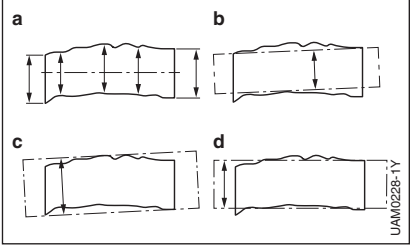
Figure 2: Independence principle
1 Envelope with maximum material virtual size.



material size forms an ideal envelope (Figure 1). All dimensional, form and parallelism deviations must be within this envelope. This requirements can be applied to cylinders and parallel, opposing surfaces. The envelope requirement was described in 1987 in DIN 7167 [2] and established this principle as standard if nothing different was indicated in the drawing.

2. The independence principle
In 1985 ISO 8015 defined the independence principle for the first time. According to this, dimensional and geometrical tolerances are independent of each other. Adding the two tolerances produces an envelope with a maximum material virtual size (Figure 2). The entry "Tolerancing acc. to ISO 8015" is absolutely necessary for drawings to which the independence principle should apply. This principle has been the GPS standard since 2012 (after DIN 7167 was withdrawn) and is also valid without a corresponding entry in the drawing. If the envelope requirement is to apply to an individual dimension element, an Ⓔ must be entered after the dimension. The independence principle can be canceled with the drawing entry "Dimensions acc. to ISO 14405 Ⓔ". In this case the envelope requirement applies to the entire drawing.

Figure 3: Dimensions acc. to ISO 14405
a) Two-point dimension (local point) LP,
b) Maximum inscribed dimension GX,
c) Minimum circumscribed dimension GN,
d) Averaged dimension (acc. to Gauss) GG.



Dimensional tolerances
Dimension definition
Although dimensions are as old as the technical era, dimension characteristics were defined for the first time in 2010 in ISO 14405 [3]. According to this, different dimension characteristics can be assigned to a dimension element, e.g., a cylinder, by indicating the corresponding modification symbol after the dimension. Four different dimension characteristics and the associated modification symbols are indicated in Figure 3. The two-point dimension (local point) does not have to be indicated since without drawing indication all the dimensions have to be measured accordingly (GPS standard).

ISO tolerance system
A special tolerance system has been defined for the pairing of two dimension elements in ISO 286 [4]. It applies to cylinders (cylindrical fits) and two parallel, opposing surfaces (flat fits). A special example of this is given in the chapter "Roller bearings" for bearing seats (see Roller bearings).

The tolerance class is identified by a combination of letters and code numbers. Capital letters are used for bores, lower-case letters for shafts. In this way the fundamental deviation, which indicates the position of the tolerance interval relative to the zero line, is defined (Figure 4). The code number (01 through 18) defines the fundamental tolerance degree (size of the tolerance). Both the fundamental deviation and the tolerance degree are indicated in ISO 286 in tables.

All combinations from A through z and from a through Z and all code numbers from 01 through 18 are possible as fits (e.g., $30^{H7}/30_{e6}$). Clearance, transition and interference fits can be defined in this way. To limit the multitude of possible fits, only bores with the fundamental deviation zero (code number H) are paired with different shaft tolerances. This produces the system of fits "Unit bore" (Figure 5). This is used whenever expensive tools (e.g., reamer) are used to produce bores. ISO 286 also defines a system of fits "Unit shaft". Only shafts with the fundamental deviation zero (code number h) are used. Clearance, transition and inter-

ference fits can be achieved with these combinations.

Geometrical tolerances

Since a two-point dimension cannot define geometry properties, it additionally requires an indication with regard to the permissible form or positional deviations.

A form tolerance only defines how far the real form may deviate from the ideal form and therefore requires no reference. A surface must for example be flat within a specified tolerance, regardless of the position of this surface.

Positional tolerances limit the permissible deviations of an element from the ideal position relative to one or more elements, the so-called reference elements. A reference is always required to define a position in space. Since references are derived from real elements, they should be as precise as possible and therefore always require a form tolerance.

Six form and eleven positional tolerances (Table 1) are defined in ISO 1101 [5]. The specification for parallelism is shown in Figure 6.

General tolerances

General tolerances are used to make drawings more simple. Standard workshop precision is sufficient for all dimensions and geometry elements which are not directly important to the function or exchangeability. It can be maintained without any special effort and applies to all the dimensions and geometries which are not individually toleranced. The general tolerances are dependent on the different manufacturing processes, e.g., machining (ISO 2768, [6], [7]), metal castings (ISO 8062-3, [8]) or plastic castings (DIN 16742, [9]).

ISO 2768 is the most frequently used general tolerance and therefore indicated

Figure 5: Unit bore

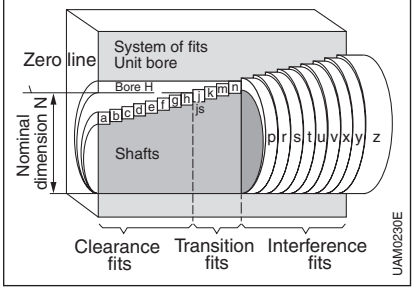
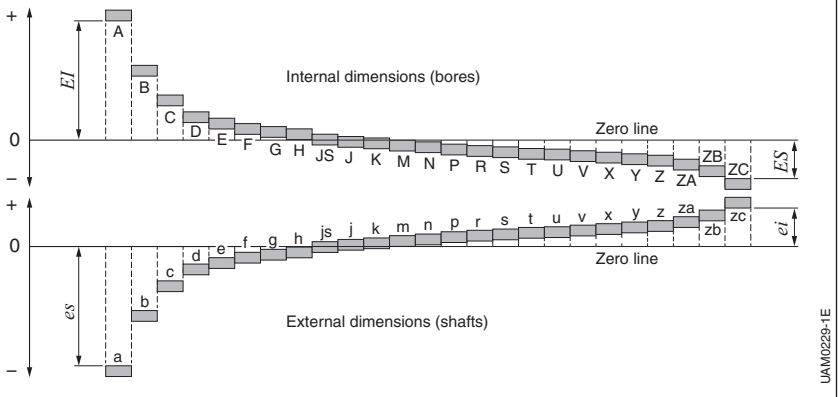


Figure 4: Position of the tolerance intervals

EI Lower limit dimension for internal dimensions, *ES* Upper limit dimension for internal dimensions,
ei Lower limit dimension for external dimensions, *es* Upper limit dimension for external dimensions.



in many drawings. The indication is provided either in or near the title block. Four tolerance classes (f, m, c, v) are available for dimensional tolerances, while only three tolerance classes (H, K, L) are available for geometrical tolerances. The following entry is required for the tolerance class "middle" (m): ISO 2768 – mH.

Surface parameters

The surface texture can dramatically influence the function of a product. Thus, the surfaces of guides and bearings should be as smooth as possible to limit wear and friction losses. However, smooth surfaces are a disadvantage to frictional joints (e.g., driving fits) since these reduce the coefficient of friction and hence also the transmittable force.

Profile meters record the profiles of the real surfaces and evaluate them according to the drawing indications. According to ISO 4287 [10] and ISO 1302 [11] the following surface parameters can be indicated and evaluated:

– primary profile, unfiltered (parameter P),
– waviness profile, without roughness profile (parameter (W),
– roughness profile, without waviness profile (parameter (R).

Primarily roughness profiles are used in practice. The parameters most commonly used in drawings are Rz and Ra (Figure 7).

Average peak-to-valley height Rz

The average peak-to-valley height Rz is the mean value of five Rz values from five adjacent individual measured lengths l_r.

$$Rz = \frac{1}{5}\sum_{i=1}^{i=5} Rz_i \ .$$

The greatest Rz value from the five individual measured lengths is designated $Rz1$max. Rt is the distance between the highest peak and the lowest valley of the profile of the overall measured length l_n. These two parameters can also be specified to describe the surface.

Arithmetic roughness average Ra

The roughness average Ra is the arithmetical mean of the absolute amounts of the deviations from the surface center line over the measured length l_n.

$$Ra = \frac{1}{l_n}\int_0^{l_n} |Z(x)|\, dx.$$

Ra was the standard characteristic value for surface texture until the 1990s. But since Ra reacts with a lack of sensitivity in respect of peaks and furrows, its informative value is minimal. In other words, Ra

Figure 6: Specification of parallelism

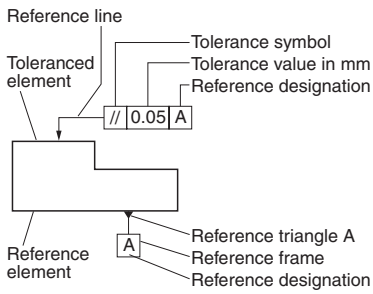


Table 1: Symbols for form and positional tolerances acc. to ISO 1101

Form tolerances without reference		Positional tolerances with reference					
		Direction		Location		Path	
—	Straightness	//	Parallelism	⌖	Position	⌒	Path
▱	Flatness	⊥	Perpendicularity	◎	Concentricity	⌒↗	Overall path
○	Roundness	∠	Slope	◎	Coaxiality		
⌭	Cylindricity			⌯	Symmetry		
⌒	Line form			⌒	Position of line		
⌓	Area form			⌓	Position of an area		

UAM0231Y

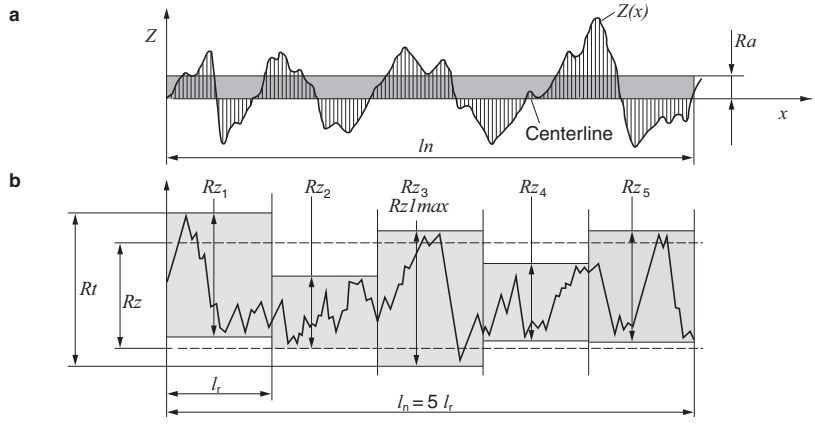
filters very strongly so that for example a single scratch is not evaluated. This is the reason why Ra is used increasingly less and Rz increasingly more for this purpose.

References

[1] DIN EN ISO 8015: Geometrical product specifications (GPS) – Basic principles – Concepts, principles and rules (ISO 8015:2011); German version EN ISO 8015:2011.
[2] DIN 7167: Relationship between tolerances of size, form and parallelism; envelope requirement without individual indication on the drawing.
[3] ISO 14405: Geometrical product specifications (GPS) – Dimensional toleranc-ing.
Part 1: Linear sizes.
Part 2: Dimensions other than linear sizes.
Part 3: Angular sizes.
[4] ISO 286: Geometrical product specifications (GPS) – ISO code system for tolerances on linear sizes.
[5] ISO 1101: Geometrical product specifications (GPS) – Geometrical tolerancing – Tolerances of form, orientation, location and run-out.
[6] DIN ISO 2768-1: General tolerances; tolerances for linear and angular dimensions without individual tolerance indications.
[7] DIN ISO 2768-2: General tolerances; geometrical tolerances for features without individual tolerance indications.
[8] ISO 8062: Geometrical product specifications (GPS) – Dimensional and geometrical tolerances for molded parts.
Part 1: Terms.
Part 3: General dimensional and geometrical tolerances and machining allowances for castings.
[9] DIN 16742: Plastics molded parts – Tolerances and acceptance conditions.
[10] ISO 4287: Geometrical Product Specifications (GPS) – Surface texture: Profile method – Terms, definitions and surface texture parameters.
[11] ISO 1302: Geometrical Product Specifications (GPS) – Indication of surface texture in technical product documentation.

Figure 7: Surface parameters

a) Arithmetic roughness average Ra,
b) Average peak-to-valley height Rz.
Rt Distance between highest peak and lowest valley of the profile of the overall measured length l_n,
l_r Individual measured length, l_n Overall measured length,
$Rz_1 \ldots Rz_5$ Average peak-to-valley height of the individual measured lengths,
$Rz1max$ Greatest Rz value from the five individual measured lengths,
Z Absolute amount of the deviations from the surface center line.



Axles and shafts

Function and application

The purpose of axles and shafts is to absorb forces and torques and transmit them to the casing. But they differ with regard to their function. Thus, an axle never transmits a torque, but a shaft always transmits a torque. While an axle can be stationary or rotate, a shaft is always rotatory. While most shafts rotate (e.g., crankshaft), they can however also perform smaller angular movements (e.g., slewing drives).

Axles
Axle serve exclusively to accommodate wheels, rollers, drums, etc. Rotating machine parts are seated on stationary axles. A rotating axle rotates in bearings about itself, while for example a wheel is permanently mounted on it. Figure 1a shows a fixed axle customary in vehicle manufacturing, while Figure 1b features a rotating axle for a railroad car. The purpose of axles is therefore confined to bearing and supporting. In practice, however, this definition is not always applied consistently, as the example of final drive shows (see Final drive).

Shafts
Shafts must not only accommodate elements and support forces, but also transmit power. Power P is calculated with torque M_t and angular frequency ω or rotational speed n as

$$P = M_t\,\omega = M_t \cdot 2\pi\, n.$$

Machine parts such as gears, belt pulleys, clutch hubs and many more are used for input and output. Shafts can be found in every rotating drive system, e.g., the crankshaft in the engine (see Crankshaft) or a shaft in the transmission (Figure 2).

Dimensioning

It is possible to observe in axles and shaft three failure causes, according to which they must then be designed and configured:
– exceeding the load capacity,
– impermissible deformation,
– operation in the resonant range (dynamic response).

Figure 1: Axles
a) Fixed vehicle axle,
b) Rotating railroad axle.
1 Axle, 2 Wheel, 3 Bearing.
F_{Whl} Wheel force, F_{Spring} Spring force,
F_{Brng} Bearing force.

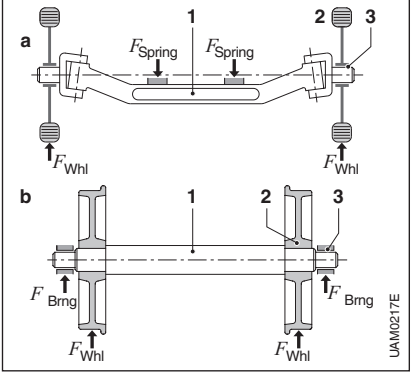
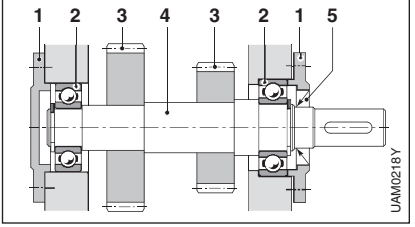


Figure 2: Transmission shaft with locating/ floating bearing arrangement
1 Bearing cap, 2 Roller bearing,
3 Gears, 4 Shaft,
5 Shaft seal.



Load capacity
Axles and shafts fail when the existing stresses are greater than the sustainable stresses. The load cases are important as well as the magnitude of the stresses. In other words, whether the stress involved is static or dynamic.

The essential differences between axles and shafts with regard to stresses are set out in Table 1. It can be seen from this that the bending stress is always alternating for axles and shafts. An exception is the eccentric shaft, where the centrifugal force rotates as a result of imbalance with the shaft and hence gives rise to a static bending stress.

In the case of a stationary axle, the bending stress can, depending on the application, be steady, cyclic or alternating.

Shafts are additionally subjected by the torque to torsional stress. Because bending and torsional stresses occur at the same time, a strength hypothesis is required to calculate the existing stress since these stresses cannot simply be summarized by addition. A precise re-calculation based on the state of the art can be performed in accordance with DIN 743 [1]. The results are for the most part good, but the work involved is correspondingly high. With less work the load capacity of a shaft can be roughly calculated with the aid of the yield criterion [2].

Table 1: Stresses

	Movement	Stress	Load case
Axle	stationary	Bending	steady, cyclic, alternating[1]
	rotating	Bending	alternating
Shaft	rotating	Bending	alternating
		Torsion	steady, cyclic, alternating[1]
Eccentric shaft	rotating	Bending	steady
		Torsion	steady, cyclic, alternating[1]

[1] Depending on the application.

Deformation
If the permissible deformation is exceeded, axles and shaft can impair the function even when the load capacity is sufficient. Transverse forces give rise to deflections, to which bearings and gears react with great sensitivity (Figure 3). Both the edge loads on the bearings and the engagement faults at the gears generate noises and the service life is greatly reduced as a result. The maximum deflection can be easily calculated for a constant cross-section. However, axles and shafts usually have complex geometries such that deflection lines are determined by means of FEM calculations.

Shafts are additionally twisted by the torsional moment, which can have a negative impact on the transmission behavior of shaft/hub connections. Therefore the torsion angle φ should not exceed 0.25° per meter length. The two ends of a shaft section with constant shaft cross-section at distance l are twisted by the torque M_t through the angle φ:

$$\varphi = \frac{M_t l}{G I_p} \cdot \frac{180°}{\pi}.$$

G is the shear modulus and can be taken for steel with sufficient accuracy as 80 GPa. I_p is the polar planar moment of inertia, for solid shafts with diameter d the following applies:

$$I_p = \frac{\pi}{64} \cdot d^4.$$

Dynamic response
Shafts are elastic components which have mass and are additionally subjected to masses such as gears, belt pulleys, etc. Therefore rotating axles and shafts are oscillating systems which are excited by centrifugal forces or by rhythmic force or torque fluctuations to forced oscillations. Resonance occurs when the excitation frequency coincides with the natural frequency. The amplitudes are in this case theoretically of infinite magnitude. In other words, an axle or a shaft can already be destroyed under very small loads if the system has enough time to oscillate in the resonant range.

Design

Axles and shafts should be able to be produced cost-effectively. In terms of design, this means providing for the fewest possible machined surfaces, using large tolerances, and permitting large peak-to-valley heights. As well as low manufacturing costs, operational reliability must also be guaranteed. The design requirements result from the functions to be performed and the mountability, the required load capacity, the permissible deformation, and the dynamic response.

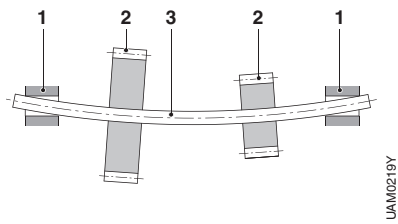
Function and mountability
A shaft as pictured in Figure 2 must perform a number of functions: directing torque via shaft/hub connections, supporting forces via bearings, preventing axial displacement of the bearing rings, defining axial positions of the gears, and sealing lubricant. A seal makes other demands on the shaft surface than a bearing seat or a shaft/hub connection. It follows from this that as a rule each function has a separate shaft diameter. Securing the axial position of bearing inner rings and hubs can be performed by means of shaft shoulders and retaining rings. When installing the bearings it is important to ensure that no sealing surfaces are damaged. This necessarily results in a stepped shaft with different diameters.

Load capacity
Shaft shoulders, which are required to ensure the function, give rise to stress concentrations as a result of the notch effect and thereby reduce the load capacity. Notches are also created by parallel keys and grooves for retaining rings. Since rotating axles and shafts are dynamically stressed by rotating bending, they are particularly notch-sensitive. Notches must therefore be avoided in highly stressed zones. If this is not possible, the notch effect must be reduced as much as possible by design measures such as small shaft shoulders and larger radii on the inside edge.

Deformation
Deformability is essentially dependent on the moment of resistance. Rigid axles and shafts with large moments of resistance are required for small deformations. It is to be noted here that hollow shafts have much better material utilization than solid shafts. Weight savings of up to 50 % are readily possible.

Dynamic response
High torsional and bending stiffness result in high natural frequencies. Therefore rotating axles and shafts are designed to be as stiff as possible so that the resonant speed is sufficiently removed from the operating speed. As with deformations, hollow shafts here too present the advantage that they offer maximum stiffness with minimum mass.

Figure 3: Edge load and misalignment due to deflection
1 Bearings,
2 Gears,
3 Shaft.



[1] DIN 743: Calculation of load capacity of shafts and axles.
Part 1: Basic principles.
Part 2: Theoretical stress concentration factors and fatigue notch factors.
Part 3: Strength of materials.
Part 4: Fatigue limit, endurance limit – Equivalently damaging continuous stress.
[2] Haberhauer/Bodenstein: Maschinenelemente. 18th Edition, Verlag Springer Vieweg, 2017.

Springs

Basic principles

Functions

All elastic components to which forces are applied are spring elements. However, springs in the narrower sense mean only those elastic elements that can absorb, store, and release work over a relatively long distance. The stored energy can also be used to maintain a force. The most important applications of industrial springs are:

- Absorbing and damping shocks (shock absorbers and vibration dampers),
- Storing potential energy (spring motors),
- Applying a force (coupling springs),
- Vibrating systems (vibrating table),
- Force measurement (spring balance).

Spring characteristic

The spring characteristic shows the behavior of a spring or spring system. This means the dependency of spring force F or spring torque M_t on deformation. Metal springs have linear characteristics (Hooke's law), rubber springs have progressive characteristics, and disc springs have degressive characteristics (Figure 1a). The gradient of the characteristic is called the spring rate R.

For translational motion: $R = \frac{dF}{ds}$.

For rotary motion: $R_t = \frac{dM_t}{d\alpha}$.

Spring duty

For frictionless springs under stress, the area under the characteristic represents the absorbed or released work (Figure 1b):

$$W = \int F\, ds\ .$$

Spring damping

If friction occurs, the prevailing force when the spring is loaded is greater than when the load is removed. The area enclosed by the two characteristics represents frictional work W_R, and is thus a measure of the damping rate (Figure 1b):

$$\psi = \frac{W_R}{W}\ .$$

Damping due to internal friction can be very high with rubber springs ($0.5 < \psi < 3$). With metal springs, however, it is rather low ($0 < \psi < 0.4$). This means that metal springs have a notable damping rate that is only achievable by means of external friction, e.g. as occurs in layers of leaf and disc springs.

Table 1: Symbols and units

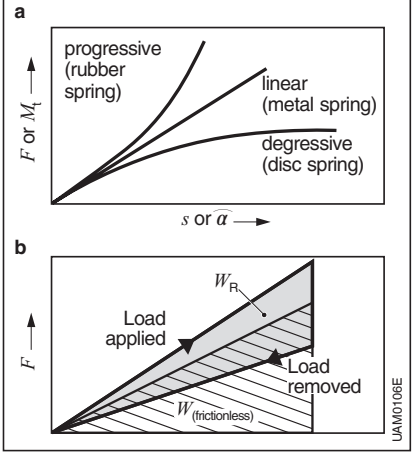
Quantity		Unit
b	Width of spring leaf	mm
d	Wire diameter	mm
D	Mean coil diameter	mm
E	Modulus of elasticity	MPa
F	Spring force	N
G	Shear modulus	MPa
h	Height of spring leaf	mm
h_0	Spring deflection (disc spring)	mm
i	Number of leaves (leaf spring)	–
i'	Number of leaves which continue to ends	–
k	Stress coefficient	–
L_c	Block length (solid length)	mm
l_f	Active length	mm
M_b	Bending moment (spring torque)	Nm
M_t	Torsional moment (spring torque)	Nm

Quantity		Unit
n	Number of active coils	–
n_t	Total number of coils	–
R	Spring rate (spring constant)	N/mm
R_t	Torsional spring rate	Nm/rad
s	Spring deflection	mm
S_a	Total of minimum distances	mm
t	Thickness (disc spring)	mm
W	Spring duty	J
W_R	Frictional work	J
$\widehat{\alpha}$	Twist angle	rad
σ_a	Permissible variable stress	MPa
σ_b	Bending stress	MPa
σ_m	Mean stress	MPa
τ_t	Torsional stress	MPa
ψ	Damping	–

Spring combinations

A very wide variety of spring characteristics can be achieved by combining several springs (Figure 2). In principle, springs can be combined in parallel or in series. A combination of parallel and series springs is also possible.

Figure 1: Spring characteristics and spring duty
a) Spring characteristics of different springs,
b) Spring duty.
s Spring deflection, $\widehat{\alpha}$ Twist angle,
F Spring force, M_t Spring torque,
W Spring duty,
W_R Frictional work.



Parallel combinations

If springs are arranged in parallel, the external load is distributed proportionally between the individual springs. However, the spring deflection (s) is equal for all springs. The spring rate of the spring system is the sum of individual spring rates:

$$R_{total} = R_1 + R_2 + R_3 + ... + R_n .$$

Accordingly, spring systems comprising parallel springs are harder than individual springs.

Series combinations

With series springs, the total external load acts on each individual spring. However, spring travel for each spring is different depending on the individual spring rates, and are added. The following applies to the resulting spring rate of the overall system:

$$\frac{1}{R_{total}} = \frac{1}{R_1} + \frac{1}{R_2} + ... + \frac{1}{R_n} .$$

Spring systems consisting of series springs are softer than the softest individual springs.

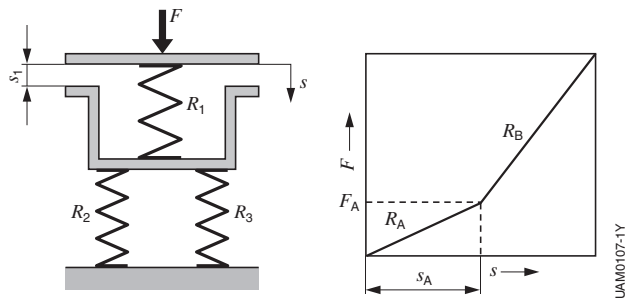
Figure 2: Spring combinations
Parallel combination of spring 2 and spring 3 in series with spring 1.

$$R_B = R_2 + R_3$$

$$\frac{1}{R_A} = \frac{1}{R_1} + \frac{1}{(R_2 + R_3)}$$

$$F_A = s_1 R_1$$

$$s_A = s_1 + \frac{F_A}{R_B}$$



Metal springs

Normally, metal springs are classified according to their stresses (Table 2). It should be noted that a loss of power occurs under permanent stress and at higher temperatures; this loss of power is called relaxation. This loss of power, which increases with rising temperature and load duration can be taken from material-dependent relaxation diagrams for compression springs in DIN EN 13906-1 [1]. From 120 °C, relaxation is no longer negligible. In unalloyed spring steels a loss of power can already occur from 40 °C.

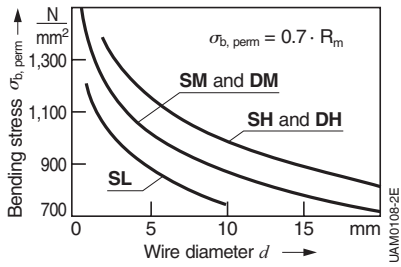
Table 2: Stress on springs

Stress	Construction
Tensile, compressive stress	Tensile test bar, ring spring
Bending stress	Leaf spring, torsion spring, spiral spring, disc spring
Torsional stress	Torsion-bar spring, helical spring

Table 3: Permissible bending stresses for leaf springs

Steel bands	Static stress $\sigma_{b,\,perm}$	Dynamic stress $\sigma_{b,perm}=\sigma_m \pm \sigma_A$
Hot-rolled	960 MPa	
Cold-rolled, hardened, and draw-tempered	1000 MPa	
Individual leaves ground		500 ± 320 MPa
Individual leaves with rolling skin		500 ± 100 MPa
Layered leaves with rolling skin		500 ± 80 MPa

Figure 3: Permissible bending stresses
R_m Tensile strength.



Springs subjected to tensile and compression stress

Due to their high spring stiffness (high spring rate), metal tensile and compression test bars are suitable only for very few special applications.

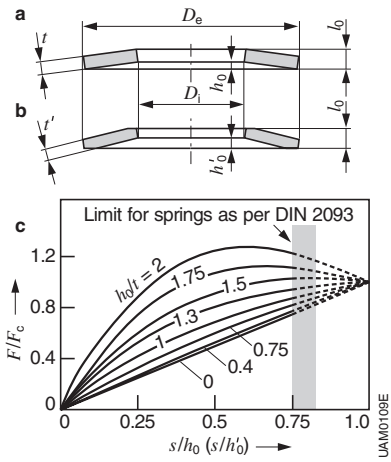
Springs subjected to bending stress

The calculation equations for leaf, coil and spiral springs subjected to bending stress are set out in Table 4.

Leaf springs

A simple leaf spring is used as a compression or guide spring. Layered leaf springs are used for suspension and wheel control in vehicles. They are usually made of spring steel in accordance with DIN EN 10089 [2] (hot-rolled) and DIN EN 10132 [3] (cold-rolled strip). The draft design may assume the permissible bending stresses specified in Table 3.

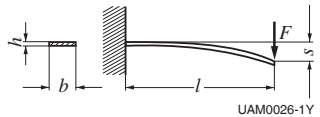
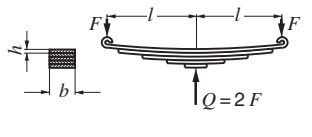
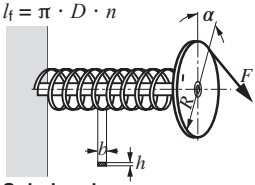
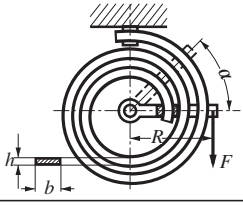
Figure 4: Disc springs according to DIN 2093
a) Without contact bearing surfaces,
b) With contact bearing surfaces,
c) Calculated spring characteristic of a disc spring to DIN 2092.



Torsion and spiral springs
During an excursion of the torsion or spiral spring (twist angle $\widehat{\alpha}$) aligning torques are generated about the rotational axis. Due to the clamping conditions, the bending stresses in the angle are almost uniform. The calculation equations for circular and rectangular sections, which apply to both torsion and spiral springs, are set out in Table 4. For the frequently used spring-steel grades (SL, SM or DM and SH or DH) it is possible, for design purposes in the case of static and quasi-static stress, when the increase in stress as a result of the wire curvature is disregarded, to calculate with the permissible bending stresses in accordance with Figure 3.

Disc springs
The cone-bowl-shaped disc springs (Figure 4) are primarily subjected to bending stresses. A wide variety of applications results from the large number of possible parallel and series combinations. Disc springs are mainly used where spring force and travel must be absorbed within confined spaces. They are used among other things in regular and

Table 4: Leaf spring, torsion spring, and spiral spring

Type		Spring force	Spring deflection
Simple straight leaf spring (constant cross-section)  UAM0026-1Y		$F_{max} = \frac{b \cdot h^2}{6 \cdot l} \cdot \sigma_{b,perm}$	$s = \frac{4 \cdot F \cdot l^3}{E \cdot b \cdot h^3}$
		Spring rate	Spring duty
		$R = \frac{F}{s} = \frac{E \cdot b \cdot h^3}{4 \cdot l^3}$	$W_{max} = \frac{\sigma^2_{b,perm}}{18 \cdot E} \cdot b \cdot h \cdot l$
Layered leaf spring  UAM0028-1Y		Spring force, spring torque	Spring deflection
		$F_{max} = \frac{i \cdot b \cdot h^2}{6 \cdot l} \cdot \sigma_{b,perm}$	$s = \frac{12 \cdot F \cdot l^3}{(2 \cdot i + i') \cdot E \cdot b \cdot h^3}$
		Spring rate	Spring duty
		$R = \frac{(2i + i') \cdot E \cdot h^3 \cdot b}{12 \cdot l^3}$	$W_{max} = \frac{\sigma^2_{b,perm}}{3 \cdot E} \cdot \frac{i^2}{(2i + i')} \cdot b \cdot h \cdot l$
Torsion spring (leg spring) $l_f = \pi \cdot D \cdot n$  UAM0029-1Y	Circular section	Spring force, spring torque	Spring deflection
		$M_{t,max} = \frac{\pi \cdot d^3}{32} \cdot \sigma_{b,perm}$	$\widehat{\alpha} = \frac{64 \cdot M_t \cdot l}{E \cdot \pi \cdot d^4}$
		Spring rate	Spring duty
		$R = \frac{M_t}{\widehat{\alpha}} = \frac{E \cdot \pi \cdot d^4}{64 \cdot l}$	$W_{max} = \frac{\sigma^2_{b,perm}}{32 \cdot E} \cdot \pi \cdot d^2 \cdot l$
Spiral spring $l_f = 2 \cdot \pi \cdot n \cdot \lvert r_0 + 0.5 \cdot n \cdot (h + \delta_r) \rvert$  UAM0030-1Y	Rectangular section	Spring force, spring torque	Spring deflection
		$M_{t,max} = \frac{b \cdot h^2}{6} \cdot \sigma_{b,perm}$	$\widehat{\alpha} = \frac{12 \cdot M_t \cdot l}{E \cdot b \cdot h^3}$
		Spring rate	Spring duty
		$R = \frac{M_t}{\widehat{\alpha}} = \frac{E \cdot b \cdot h^3}{12 \cdot l}$	$W_{max} = \frac{\sigma^2_{b,perm}}{6 \cdot E} \cdot b \cdot h \cdot l$

overload clutches and to pre-tension rolling bearings. Because friction occurs between the individual disc springs in spring assemblies layered in the same direction, they are also suitable for damping vibrations and shocks.

At h_0/t > 0.4, the nonlinearities of the springs are no longer negligible. Spring forces, spring travel, and spring rates can be calculated with sufficient precision according to DIN 2092 [4], or they can be taken from the manufacturer's data.

For disc springs subjected to static stress (< 10^4 stress reversals), fatigue-strength calculation is not required if the maximum spring force at $s = 0.75\ h_0$ is not exceeded. For dynamically stressed disc springs the maximum permissible tensile stress and the permissible stress range can be taken from DIN 2093 [5].

Springs subjected to torsional stress

The calculation equations for torsion-bar and helical springs subjected to torsional stress are set out in Table 5.

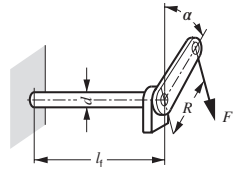
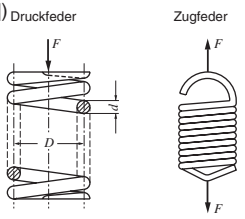
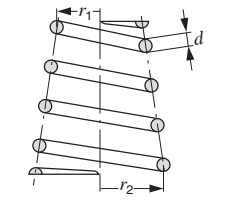
Torsion-bar springs

Circular sections are usually selected for torsion bars. They have a very high volume utilization factor, which means that they can absorb a lot of energy, but they occupy little space.

Helical springs

Cylindrical helical springs are manufactured as compression and extension

Table 5: Torsion-bar spring and helical springs

Type	Spring force, spring torque	Spring deflection
Torsion-bar spring Circular section (DIN 2091, [6]) 	$M_{t,\,max} = \frac{\pi \cdot d^3}{16} \cdot \tau_{t,\,perm}$	$\widehat{\alpha} = \frac{32 \cdot M_t \cdot l_f}{G \cdot \pi \cdot d^4}$
	Spring rate	Spring duty
	$R = \frac{M_t}{\widehat{\alpha}} = \frac{G \cdot \pi \cdot d^4}{32 \cdot l_f}$	$W_{max} = \frac{\tau^2_{t,perm}}{16 \cdot G} \cdot \pi \cdot d^2 \cdot l_f$
Cylindrical helical springs Circular section (DIN EN 13906, [1]) 	Spring force, spring torque	Spring deflection
	$F_{max} = \frac{\pi \cdot d^3}{8 \cdot k \cdot D} \cdot \tau_{t,\,perm}$	$s = \frac{8 \cdot D^3 \cdot n}{G \cdot d^4} \cdot F$
	Spring rate	Spring duty
	$R = \frac{G \cdot d^4}{8 \cdot D^3 \cdot n}$	$W_{max} = \frac{\tau^2_{t,perm}}{16 \cdot G} \cdot d^2 \cdot D \cdot \pi^2 \cdot n$
Tapered helical springs with circular section 	Spring force, spring torque	Spring deflection
	$F_{max} = \frac{\pi \cdot d^3}{16 \cdot k \cdot r_2} \cdot \tau_{t,\,perm}$	$s = \frac{16 \cdot (r_1 + r_2) \cdot (r_1^2 + r_2^2) \cdot n \cdot F}{G \cdot d^4}$
	Spring rate	Spring duty
	$R = \frac{G \cdot d^4}{16 (r_1 + r_2) \cdot (r_1^2 + r_2^2) \cdot n}$	$W_{max} = \frac{\tau^2_{t,perm}}{32 \cdot G} \cdot \frac{d^2 (r_1 + r_2) \cdot (r_1^2 + r_2^2) \cdot \pi \cdot n}{r_2^2}$

springs. The calculation equations are identical for both types. Compression springs in conical form can optimize the use of space if the individual coils can be pushed into each other.

Compression springs

To enable the force to be introduced as centrically as possible in the case of a compression spring, the spring ends are created and surface-ground vertically to the spring axis. Cold-formed springs consist of at least two active coils ($n \geq 2$) and two created, non-active end coils. In the case of hot-formed springs n should be ≥ 3 and as non-active end coils only ¾ coils are created in each case. This produces the total number of active coils as indicated in Table 6. Furthermore, the spring ends are arranged with an offset of 180° to each other so as to produce a total coil number of n_t = 2.5; 3.5; 4.5; 5.5; 6.5, etc.

In order to avoid plastic deformation in the spring, the active coils may not touch each under maximum load. In other words, a compression spring must never be compressed to the block (solid) length

Figure 6: Permissible torsional stresses for helical springs with static stress

a) Cold-formed from patented drawn spring-steel wires (SL, SM, DM, SH and DH) and valve spring-steel wire (VDC) acc. to DIN EN 10270-2,
b) Hot-formed from spring steels acc. to DIN EN 10089.
R_m Tensile strength.

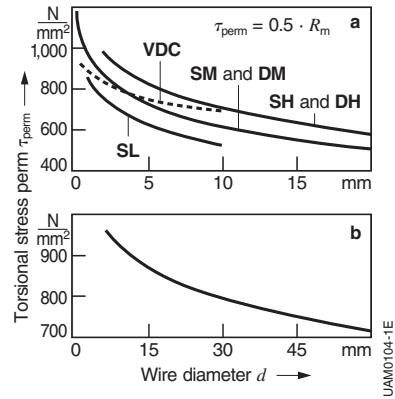
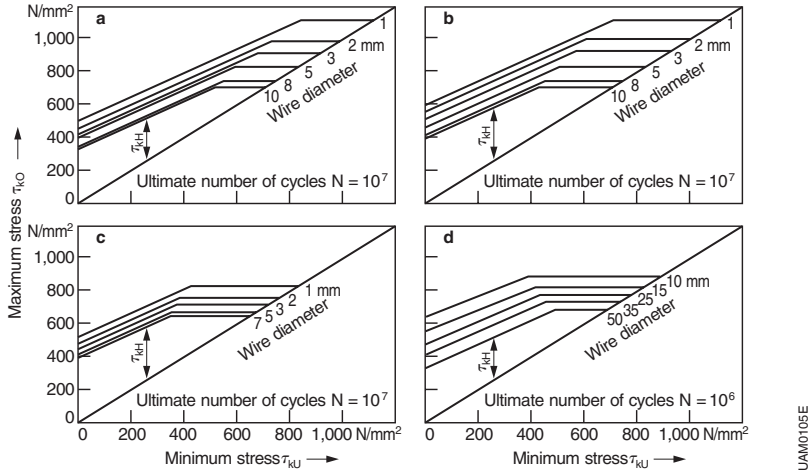


Figure 5: Fatigue-limit diagrams for helical compression springs

a) For cold-formed springs made from spring-steel wires SH and DH (not shot-blasted),
b) For cold-formed springs made from spring-steel wires SH and DH (shot-blasted),
c) For cold-formed springs made from valve spring-steel wire (VDC),
d) For hot-formed springs.



L_c. The sum of the minimum distances between the individual coils is indicated with S_a and can be calculated in accordance with Table 6. Under dynamic load S_a must be doubled for hot-formed springs, whereas a factor of 1.5 should be used for cold-formed springs.

The effect of the wire curvature is taken into account by the stress coefficient k (Table 7). In the case of static stress, this effect can be disregarded, i.e. k = 1 is then set. The following applies to the stress range prevailing in the case of dynamic stress:

$$\tau_{kh} = k \frac{8\,D}{\pi \cdot d^3} \cdot (F_2 - F_1) \leq \tau_{kH} .$$

Extension springs

Extension springs are formed either with loops or with rolled-in or screwed-in end pieces. Since service life is determined primarily by the loops, it is impossible to give general fatigue limit values. Cold-formed extension springs hardened and tempered after drawing can be manufactured with an internal preload. This allows significantly higher stresses.

References

[1] DIN EN 13906: Cylindrical helical springs made from round wire and bar – Calculation and design. Part 1: Compression springs.
Part 2: Extension springs.
Part 3: Torsion springs.
[2] DIN EN 10089: Hot rolled steels for quenched and tempered springs – Technical delivery conditions.
[3] DIN EN 10132: Cold-rolled narrow steel strip for heat-treatment – Technical delivery conditions.
Part 1: General.
Part 2: Case hardening steels.
Part 3: Steels for quenching and tempering.
Part 4: Spring steels and other applications.
[4] DIN 2092: Disc springs – Calculation.
[5] DIN 2093: Disc springs – Quality requirements – Dimensions.
[6] DIN 2091: Circular section torsion bar springs; calculation and design.
[7] DIN-Taschenbuch 29: Federn 1. Berechnungs- und Konstruktionsgrundlagen, Qualitätsanforderungen, Bestellangaben, Begriffe, Formelzeichen und Darstellungen. Beuth-Verlag 2015.
[8] DIN-Taschenbuch 349: Federn 2. A compilation of the current material standards and definitions for semi-finished products. Beuth-Verlag 2012.
[9] H. Haberhauer; F. Bodenstein: Maschinenelemente. 18th Edition, Verlag Springer Vieweg, 2017.
[10] F. Fischer; H. Vondracek: Warm geformte Federn – Konstruktion und Fertigung. Hohenlimburg – Hoesch AG, 1987.
[11] M. Meissner; H.-J. Schorcht; U. Kletzin: Metallfedern – Grundlagen, Werkstoffe, Berechnung, Gestaltung und Rechnereinsatz. 3rd Ed., Verlag Springer Vieweg, 2015.

Table 6: Helical springs

	Total number of coils	Block length (solid length)	Sum of minimum distances
Cold-formed	$n_t = n + 2$	$L_c \leq n_t \cdot d$	$S_a = (0.0015 \cdot D^2/d + 0.1 \cdot d) \cdot n$
Hot-formed	$n_t = n + 1.5$	$L_c \leq (n_t - 0.3) \cdot d$	$S_a = 0.02 \cdot (D + d) \cdot n$

Table 7: k-factor

D/d	3	4	6	8	10	14	20
k	1.55	1.38	1.24	1.17	1.13	1.10	1.07

Friction bearings

Features

The task of bearings is to guide machine parts and support forces between surfaces that are moving relative to each other. In order to minimize the wear and power loss, the friction that occurs in the process must be kept as low as possible.

Friction (or plain or sliding) bearings are made of metallic (e.g. sintered metals) and non-metallic materials. A distinction is made between radial and axial (thrust) bearings based on the direction of the external load. Another distinguishing feature is lubrication. Hence, friction bearings can be either hydrostatically or hydrodynamically lubricated, or self-lubricating. A complete separation of the sliding surfaces is the optimal state. With full lubrication there is pure fluid friction with very low coefficients of friction. The lubricant layer has a shock-, vibration- and noise-absorbing effect.

Table 1: Symbols and units (DIN 31652 [1])

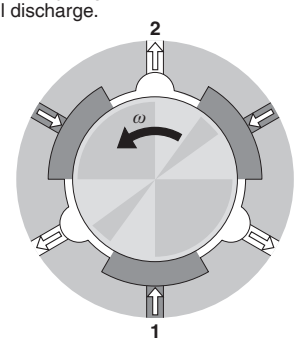
Description	Symbol	Unit
Axial bearing length	B	m
Inside bearing diameter (nominal diameter)	D	m
Shaft diameter (nominal diameter)	d	m
Eccentricity (displacement between shaft and bearing centers)	e	m
Bearing force (load)	F	N
Minimum lubricating film thickness	h_{min}	m
Local lubricating film pressure	p	Pa = N/m^2
Specific bearing load = $F/(B\,D)$	$\overline{p}$	Pa
Bearing clearance = $D - d$	s	m
Sommerfeld number	So	–
Sliding velocity	v	m/s
Relative eccentricity = $2\,e/s$	ε	–
Effective dynamic viscosity of lubricant	η_{eff}	Pa · s
Relative bearing clearance = s/D	ψ	–
Hydrodynamically effective angular velocity	ω_{eff}	s^{-1}

Hydrostatic friction bearings

In the case of hydrostatic bearings the lubricant is forced by an external pump at high pressure between the sliding surfaces. At any point in time there is a complete separation of the sliding surfaces on account of the thin lubricating film (Figure 1). Friction losses only occur on account of the shear forces in the lubricant. These forces are proportional to the speed at which the bearing surfaces move against each other. At low relative speeds hydrostatic friction bearings consequently operation with virtually zero friction. There is no stick-slip effect when starting and after-running due to the absence of slip resistance.

Hydrostatic friction bearings are therefore wear-and-tear-resistant and consequently have a very long service life. The drawbacks of these friction bearings are the high costs and the large installation space required. For these reasons they are used primarily in heavy machine construction.

Figure 1: Hydrostatic friction bearing.
1 Oil supply (high pressure),
2 Oil discharge.



Hydrodynamic friction bearings

Application

Most of the hydrodynamic friction bearings used in motor-vehicle engines are plain bearings (Figure 2) for holding the crankshaft drive, including the camshafts. They are usually designed as bearing shells with special clearance (oval clearance acc. to Figure 3).

Operating principle

In the case of hydrodynamic friction bearings the pressure is generated automatically. The following preconditions are required for this purpose:

– relative movement between the sliding surfaces,
– wedge-shaped gap,
– lubricant must adhere to the sliding surfaces,
– sufficient lubricant supply.

A friction bearing requires a bearing clearance in order to build up a lubricating gap that is capable of bearing. In the stationary state the shaft touches the bearing shell and consequently there is solid-body friction. The eccentric displacement of the shaft creates a wedge-shaped lubricating gap (Figure 3). As the rotational speed increases a lubricating film that is capable of bearing is built up as a result of the lubricant that adheres to the sliding surfaces being pumped into the wedge gap. Until the pressure has built up fully in the lubricating gap, i.e. until the sliding surfaces are completely separated, the shaft still partly touches the bearing shell. During this period the bearing is in the mixed-friction range. There is pure fluid friction only when the operating rpm exceeds the transition rpm.

Stribeck curve

The friction states in the hydrodynamic friction bearing were analyzed by Richard Stribeck. The curve named for him shows the dependence of friction on shaft speed (Figure 4, [2]).

Sommerfeld number

Depending on load and rotational speed, the eccentricity of the shafts is set within the system in such a way that the integral of the lubricant pressure of the external bearing force maintains equilibrium. The hydrodynamic pressure distribution in a convergent bearing gap is determined from the solution of Reynolds' differential equation. The integration of the pressure

Figure 2: Pressure distribution in plain bearing

e Eccentricity,
F Bearing force,
h_{min} Smallest lubricating gap,
d Shaft diameter,
D Bearing inside diameter,
p Pressure distribution,
ω Angular velocity,
S Shell center point (bearing),
W Shaft center point.

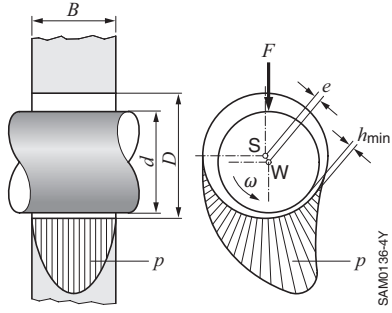
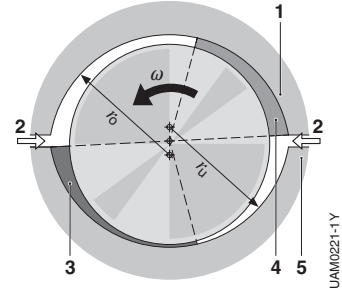


Figure 3: Two-face bearing shell with oval clearance

1 Upper bearing shell,
2 Oil supply,
3 Lower lubricating wedge,
4 Upper lubricating wedge,
5 Lower bearing shell.
r_o Radius, upper bearing shell,
r_u Radius, lower bearing shell,
ω Angular velocity.



distribution produces the load capacity of the lubricating film and is expressed in the dimensionless Sommerfeld number So [1] (for quantities, see Table 1):

$$So = \frac{F\,\psi^2}{(D\,B\,\eta_{\text{eff}}\,\omega_{\text{eff}})}.$$

For stable operation $So > 1$ must be the case, bearings subjected to high load should be designed with $So > 3$. As the Sommerfeld number increases, so the relative eccentricity ε increases and the minimum lubricating-gap thickness h_{min} decreases. The following applies:

$$h_{\text{min}} = \frac{(D-d)}{2} - e = 0.5\,D\,\psi\,(1-\varepsilon).$$

The relative eccentricity is then:

$$\varepsilon = \frac{2e}{(D-d)}.$$

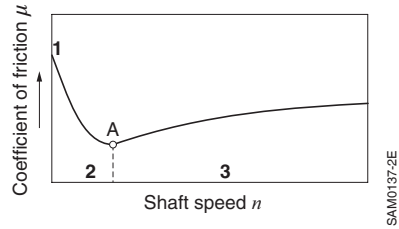
With the aid of the Sommerfeld number it is possible, in accordance with DIN 31652 [1] and VDI 2204 [2], to calculate the coefficient of friction μ in the bearing and thus also the friction loss and the thermal stress.

Table 2: Coefficients of friction for different types of friction

Type of friction	Coefficient of friction μ
Solid-body friction	0.1…>1
Mixed friction	0.01 to 0.1
Fluid friction	0.001 to 0.01

Figure 4: Stribeck curve (diagram).
1 Solid-body friction,
2 Mixed friction,
3 Fluid friction,
A Transition point (transition rpm).

Coefficient of friction $\mu = \frac{\text{Friction force}}{\text{Normal force}}$



Requirements

The coefficients of friction given in Table 2 are approximate values, and are intended only for comparison of the different operating states. As hydrodynamic bearings also operate with mixed friction some of the time, must be able to withstand a certain amount of contamination without loss of function, and are also subjected to high dynamic and thermal stress (particularly in piston engines), the bearing material must meet a number of requirements, some of which are mutually exclusive.

- Conformability: This is the property of a bearing material to adapt to necessary modifications of shape by local deformation without permanent damage.
- Embeddability: This is the ability to absorb dirt particles into the bearing surface without negative consequences.
- Resistance to wear: This prevents the separation of small particles in the event of mechanical load in the mixed-friction area.
- Run-in performance: This is the interaction of conformability, embeddability, and resistance to wear.
- Resistance to seizing: This prevents partial welding of the sliding surfaces under high compressive load.
- Mechanical loadability: This prevents plastic deformation under high compressive loads.
- Fatigue strength: This describes the slowly progressing material fatigue which under alternating load can result in failure even in the case of small bearing forces.

If a bearing (e.g. piston-pin bushing) is simultaneously subjected to high loads and low sliding velocities, high fatigue strength and wear resistance should take precedence over resistance to seizing. The bearing materials used in such cases are hard bronzes or special brass alloys.

Since they are subjected to dynamic loads with high sliding velocities, connecting-rod and crankshaft bearings in internal combustion engines must fulfill a number of different requirements. In these applications, multilayer bearings (Figure 5), above all trimetal bearings, have proved themselves in practice.

The service life of friction bearings in the crankshaft drive can be further increased through the use of special solutions such as the sputter bearing (Figure 6). It contains a highly wear-resistant AlSn running layer (sputter layer) which is applied in the PVD procedure (physical vapor deposition) to the bearing material underneath. Grooved sliding bearings (Figure 7) are also used in internal-combustion engines subjected to high load (e.g. turbocharged diesel engines). In this variant fine grooves, which are filled with a soft filler (e.g. PbSnCu), are incorporated in the circumferential direction of the bearing surface. The bearing surface therefore exhibits soft and hard areas next to each other.

Figure 5: Multilayer bearing
(Design of a trimetal bearing).
1 Steel backing shell, 2 Bearing metal,
3 Diffusion barrier (e.g. 1 – 2 µm nickel),
4 Penetration coating (approx. 20 µm, electroplated SnCu layer or anti-friction paint).

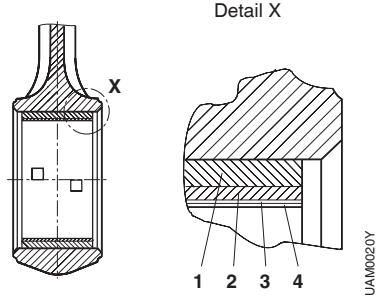
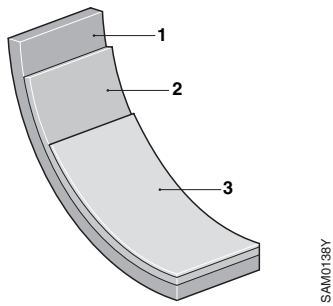


Figure 6: Section through a sputter bearing (lead-free)
1 Steel backing,
2 Intermediate layer (brass or bronze),
3 Running layer (e.g. AlSn 20 Cu).



Bearing materials

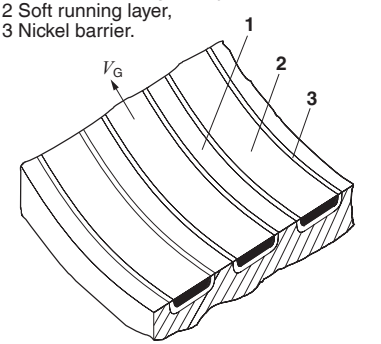
Lead, tin, copper, and aluminum alloys are used as bearing materials. Values for the permissible specific bearing load are given in Table 3.

Lead and tin bearing metals, formerly called babbitt metals, are ideally suitable for higher sliding velocities and are characterized by good run-in and emergency-running properties. Tin bronzes

Table 3: Empirical values for permissible specific bearing load (acc. to DIN 31652-3)

Bearing materials	Permissible specific bearing load $\overline{p}_{lim}$ [MPa]
Pb and Sn alloys (babbitt metals)	5 to 15
Bronze, tin base	7 to 25
Bronze, lead base	7 to 20
AlSn alloys	7 to 18
AlZn alloys	7 to 20
Maximum values only apply for very low sliding velocities	

Figure 7: Section through a grooved sliding bearing
(Miba patent).
Running surface with very fine grooves in running direction V_G.
1 Wear-resistant light alloy,
2 Soft running layer,
3 Nickel barrier.



are suitable for high stresses and are very wear-resistant; run-in and emergency-running properties are however less good. Lead bronzes demonstrate better emergency-running properties with only slightly worse resistance to wear. Aluminum alloys have a higher corrosion resistance than babbitt metal and copper alloys (bronzes).

Standardized materials for friction bearings are listed in ISO 4381 [3], ISO 4382 [4], and ISO 4383 [5]. Some bearing materials contain lead. EU 2016/774 [6] prohibits the use of lead in passenger cars. Lead is currently still permitted in commercial vehicles and in general machine construction.

Self-lubricating bearings made of metal

Sintered bearings

Friction bearings made of sintered metal belong to the category of self-lubricating bearings. They consist of sintered metals which are porous and impregnated with liquid lubricants. For many small motors in motor automotive applications, this type of bearing is a good compromise in terms of precision, installation, freedom from maintenance, service life, and cost. They are used primarily with shaft diameters from 1.5 to 12 mm.

For motor vehicles sintered-iron and sintered-steel bearings are preferred to sintered-bronze bearings since they are cheaper and exhibit a smaller interaction with the lubricant (Table 4). The advantages of sintered-bronze bearings are higher permissible sliding velocities, lower noise, and lower friction coefficients (e.g. for record-players, office equipment, and data systems).

Sintered metals bearing the designation SINT-B (Table 4) have 20 % porosity. Aside from these there are also SINT-A with 25 % and SINT C with 15 % porosity.

The performance of sintered bearings over long periods of service is closely related to the use of optimum lubricants. Mineral oils are used; however, these exhibit insufficient cold-flow properties and moderate resistance to aging. Synthetic oils (e.g. esters, poly-α-olefins) on the other hand have good cold-flow proper-

Table 4: Materials for sintered-metal bearings (acc. to DIN 30910-3 [7])

Material group	Designation Sint- …	Composition [%]	Remarks
Sintered iron	B 00	< 0.3 C < 1 Cu Rest Fe	Standard material which meets moderate load and noise requirements.
Sintered steel, containing Cu	B 10	< 0.3 C 1…5 Cu Rest Fe	Good resistance to wear, can be subjected to higher loads than pure Fe bearing.
Sintered steel, higher Cu content	B 20	< 0.2 Cu 15…20 Cu Rest Fe	Cheaper than sintered bronze, good noise behavior.
Sintered bronze	B 50	< 0.2 C 9…11 Sn Rest Cu	Standard Cu-Sn based material, good noise behavior.

ties and can be subjected to high thermal loads. Furthermore, the evaporation tendency is low. Synthetic grease oils (oils which have metal soaps) are characterized by low starting friction and low wear. The most important properties of sintered bearings are set out in Table 5.

Metal-ceramic bearings

Metal-ceramic friction bearings consist of material manufactured by powder-metallurgy processes; in addition to the metallic matrix, the bearing material also contains finely distributed solid lubricant particles. Bronze, iron, and nickel are used as the material, graphite or molybdenum disulfide (MoS_2) for example are used as the lubricant. Ceramic bearings are particularly suitable for use under high loads, and are at the same time self-lubricating. However, they are very brittle and therefore sensitive to jolts and impacts.

Metal-ceramic bearings are used in motor vehicles, for example as steering-knuckle bearings.

Table 5: Properties of maintenance-free, self-lubricating bearings

Properties / Quantity	Sintered bearings oil-impregnated: Sintered iron	Sintered bearings oil-impregnated: Sintered bronze	Polymer bearings: Thermoplastic polyamide	Polymer bearings: Duroplastic polyimide	Metal-backed composite bearings Running layer: PTFE + additive	Metal-backed composite bearings Running layer: Acetal resin	Synthetic carbons
Pressure resistance MPa	80…180		70	110	250	250	100…200
Max. sliding velocity m/s	10	20	2	8	2	3	10
Specific load MPa	1…4 (10) [1]		15	50 (at 50 °C) 10 (at 200 °C)	20…50	20…50	50
Permissible operating temperature °C	– 60…180 (depends on oil)		–130…100	–100…250	–200…280	–40…100	–200…350
Short-term	200		120	300		130	500
Coefficient of friction without lubrication	with lubrication 0.04…0.2		0.2…0.4 (100 °C) 0.4…0.6 (25 °C)	0.2…0.5 (unfilled) 0.1…0.4 (filled)	0.4…0.2	0.7…0.2 PTFE filled	0.1…0.35
Thermal conductivity W/(m·K)	20…40		0.3	0.4…1	46	2	10…65
Corrosion resistance	less good	good	very good		good	good	very good
Chemical resistance	no		very good		conditional	conditional	good
Max. $p \cdot v$ MPa·m/s	20		0.05	0.2	1.5…2		0.4…1.8
Embeddability of dirt and abraded material	less good		good	good	less good	good	less good

[1] Value in parentheses applies to additional lubrication.

Self-lubricating bearings made of plastic

Different plastics can be used for friction bearings. The various properties of plastic bearings are set out in Table 5.

Solid polymer bearings made of thermoplastics

Features

Friction bearings made of thermoplastics are inexpensive and are suitable for low bearing forces and low operating temperatures. The danger of "seizing" is extremely small.

The most frequently used thermoplastic polymer materials are:
- Polyoximethylene (POM, POMC)
- Polyamide (PA)
- Polyethylene and polybutylene terephthalate (PET, PBT)
- Polyetheretherketone (PEEK)

The tribological and mechanical properties can be varied over a wide range by incorporating lubricants and reinforcements in the thermoplastic base material.

Lubrication additives
- Polytetrafluoroethylene (PTFE)
- Graphite (C)
- Silicone oil, and other liquid lubricants, recently also enclosed in microcapsules

Reinforcement additives
- Glass fibers (GF)
- Carbon fibers (CF)

Application examples
- Windshield-wiper bearings (PA and glass fiber)
- Idle actuators (PEEK + carbon fiber, PTFE and other additives)

Polymer bearings made of duroplastics

These materials have high intrinsic friction and are seldom used as bearing materials in motor vehicles. Duroplastics which are used for friction bearings are:
- Phenolic resins (high friction)
- Epoxy resins (additives of PTFE or C, reinforcement by fibers required on account of their inherent brittleness)
- Polyimides (high thermal and mechanical loadability)

Application example

Polymer bearings made of duroplastics are used in wiper motors as axial startup blocks made of polyimide.

Metal-backed composite bearings

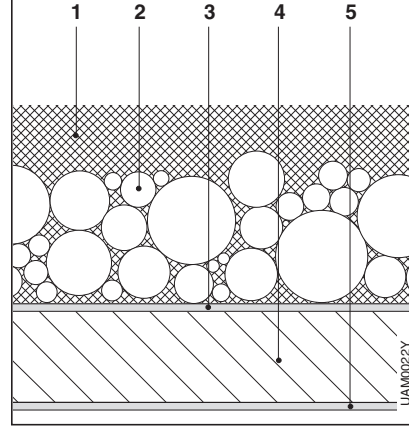
Composite bearings are combinations of polymer materials, fibers, and metals. The respective structure (Figure 8) deliver advantages over pure or filled polymer friction bearings with regard to load capacity, bearing clearance, heat conduction, and installation. They are also more suitable for oscillating motions.

Example of bearing structure

The bearing consists of a steel backing electroplated with tin; above this a

Figure 8: Section through a self-lubricating composite bearing

1 Polymer liner, 2 Porous bronze layer, 3 Copper layer, 4 Steel backing, 5 Tin layer.



0.2…0.35 mm thick bronze-bead layer with 30…40 % porosity in which a low-friction polymer material is rolled in as a liner is sintered on. The liner consists of
– acetal resin or polyvinylidene fluoride, either impregnated with oil or containing lubricating recesses, or
– PTFE + ZnS or molybdenum disulfide (MoS_2) and graphite as additive.

Metal-backed composite bearings are available in a number of different shapes and compositions. Metal-backed composite bearings with woven PTFE fiber inserts have unusually high loadability and are suitable for use in ball-and-socket joints.

Application examples for motor vehicles
– Piston-rod bearings for suspension struts
– Release-lever bearings for clutch pressure plates
– Brake-shoe bearings in drum brakes
– Ball-and-socket bearings
– Door-hinge bearings
– Bearings for winding shafts for seat belts
– Steering-knuckle bearings
– Gear-pump bearings

Composite bearings with specially modified liners are required for heavy-duty requirements in diesel high-pressure injection pumps. The liner is made of PEEK or PPS with additives (e.g. carbon fibers, ZnS, TiO_2, and graphite). The particle size is sometimes in the nanometer range.

Carbon-graphite bearings
Carbon-graphite bearings are members of the ceramic bearing family due to their method of manufacture and material properties. Powdered carbons are used as the base material, pitch and synthetic resins are used as binders. It is to be noted that carbon-graphite bearings are very brittle.

Advantages
– Thermal stability up to 350 °C (hard-burnt carbon) and up to 500 °C (electro-graphite)
– Low friction
– Good corrosion resistance
– Good thermal conductivity
– Good thermal shock resistance.

Application examples for carbon-graphite bearings
– Fuel-pump bearings
– Bearings in drying ovens
– Adjustable guide vanes of compressors in exhaust-gas turbochargers

---

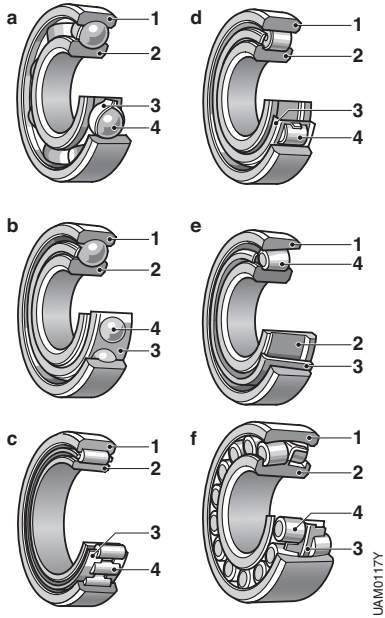
References
[1] DIN 31652: Plain bearings – Hydrodynamic plain journal bearings under steady-state conditions.
Part 1: Calculation of circular cylindrical bearings.
Part 2: Functions for calculation of circular cylindrical bearings.
Part 3: Permissible operational parameters for calculation of circular cylindrical bearings.
[2] VDI 2204: Design of plain bearings.
[3] ISO 4381: Plain bearings – Tin casting alloys for multilayer plain bearings.
[4] ISO 4382: Plain bearings; copper alloys;
Part 1: Cast copper alloys for solid and multilayer thick-walled plain bearings.
Part 2: Wrought copper alloys for solid plain bearings.
[5] ISO 4383: Plain bearings – Multilayer materials for thin-walled plain bearings.
[6] Commission Directive (EU) 2016/774 of 18 May 2016 amending Annex II to Directive 2000/53/EC of the European Parliament and of the Council on end-of-life vehicles.
[7] DIN 30910: Sintered metal materials; sintered-material specifications (WLB);
Part 1: Explanatory notes for WLB.
Part 2: Sintered metal materials for filters.
Part 3: Sintered metal materials for bearings and structural parts with bearing properties.
Part 4: Sintered metal materials for structural parts.
Part 6: Hot-forged sintered steels for structural parts.

Rolling bearings

Application

Rolling bearings are some of the most important components in machines. Great demands are made on their load capacity and operational reliability. Rolling bearings are widely used in motor vehicles. Passenger cars and commercial vehicles are fitted with a large number of rolling bearings in bearing arrangements for alternator, starter, wheel bearing, transmission, suspension strut, drive shaft, water pump, tension pulley, steering, windshield-wiper motor, fan, and fuel-injection pump.

Figure 1: Design of rolling bearings
a) Deep-groove ball bearing,
b) Angular-contact ball bearing,
c) Needle bearing,
d) Cylindrical rolling bearing,
e) Tapered rolling bearing,
f) Self-aligning rolling bearing.
1 Outer race, 2 Inner race,
3 Cage, 4 Rolling element.



General principles

Type
Rolling bearings are generally made up of two races (Figure 1, outer race and inner race), a cage, and a rolling-element assembly. The rolling elements guided by the cage roll on the raceways. Balls, cylindrical rollers, needle rollers, tapered rollers, and self-aligning rollers are used as rolling elements. A rolling bearing can be lubricated with grease. It is fitted with sheet-steel cover plates or rubber gaskets to prevent grease discharge and to provide a seal against dirt.

The rolling bearing transfers the outer force from one bearing race to the other via the rolling elements. A distinction is made between radial bearings and axial (thrust) bearings, depending on the main direction of load.

Structural dimensions
The rolling bearing is a ready-to-install machine part. Various ranges of outside diameters and widths are available for a hole diameter. Standardized abbreviation codes are used to identify the diameter and width ranges of a rolling bearing. The designations are recorded in the standard DIN 623-1 [1].

The external dimensions are indicated in the catalogs of rolling-bearing manufacturers.

Tolerances
Dimensional and form tolerances of rolling bearings are standardized in ISO 492 [2] and DIN 620 ([3], [4], [5], [6], [7]) according to precision. Rolling bearings of normal precision, i.e. tolerance class P0 (also called PN), generally satisfy all the demands which are made by mechanical engineering on bearing quality. For more stringent requirements, the standard provides for more precise tolerance classes P6, P5, P4, and P2.

Bearing play and bearing clearance
The bearing play of an uninstalled rolling bearing refers to the distance through which the bearing races can be moved against each other. A distance is made between radial play and axial play.

Radial play is defined in the standard DIN 620 Part 4 [6]. The normal radial play category is CN. In accordance with operating conditions – e.g. fits with conversion parts and temperatures – it is possible also to use the other radial-play categories C2 (< CN) or C3 and C4 (> CN).

Axial play is derived from the radial play and raceway and rolling-element geometries, and is always given as a reference parameter.

When bearings are installed the term used is bearing clearance. During operation the bearing clearance is determined by the original bearing play, fits and material of the conversion parts such as shaft and housing, and the temperature difference between the bearing races. As a rule a perfectly running bearing should have a very small bearing clearance.

Materials
Bearing races and rolling elements are primarily composed of chrome-alloy special steel 100Cr6 (DIN EN ISO 683-17, [8]) or 52100 (ASTM A295, [9]) with a high degree of purity and hardness in the range of 58 to 65 HRC. For special applications the races and rolling elements can be made of other materials, e.g. ceramic.

Rolling-bearing cages are made of metal or plastic. The metallic cage in small rolling bearings is primarily composed of sheet steel. Polyamide 66 (PA66) is used for the majority of plastic cages. This material, especially when reinforced with fiberglass, is characterized by its favorable combination of strength and elasticity.

Other thermoplastics and duroplastics are also used as cage materials for special applications subject to extremely high thermal loads.

Selection of rolling bearings

It is necessary to take into account many external factors in order to choose the right bearing from the wealth of options.

Selection criteria
Load
The size and direction of the load acting on the rolling bearing normally determine the type and size of the bearing. Deep-groove ball bearings are usually used in the case of low to medium loads. Rolling bearings have one advantage in the case of high loads and limited installation space. With the exception of only purely radially loaded needle bearings, cylindrical rolling bearings and axial bearings, rolling bearings can simultaneously accommodate radial and axial loads (combined loads).

Deep-groove ball bearings transfer axial loads in both directions, while angular-contact ball bearings and tapered rolling bearings can only be axially loaded in one direction.

Cylindrical rolling bearings and self-aligning bearings are particularly suitable for radial loads, but less so for axial loads.

Speed limit
Ball bearings with point contact between rolling elements and raceways have a higher speed limit than rolling bearings of the same size. The permissible speed of a rolling bearing is dependent above all on the type of construction, the size, and also the lubricating process. A bearing lubricated with oil usually has a higher speed limit than a bearing lubricated with grease.

Mounting
A distinction is made between locking bearings (completely assembled) and non-locking bearings (can be disassembled). Non-locking bearings include tapered rolling bearings, angular-contact ball bearings, cylindrical rolling bearings, and needle bearings. These bearings are for the most part easier to assemble and disassemble than locking bearings such as deep-groove ball bearings and self-aligning bearings. During mounting on the shaft and in the housing, tapered rolling bearings and angular-contact ball

bearings must be adjusted to the required bearing clearance and preload, which always calls for great care.

Further selection criteria
In addition to the previously mentioned main criteria, it is also necessary to take into account – when choosing a rolling bearing – angular adjustability to compensate for misalignment between the bearing points, running smoothness, friction, and costs.

Layout of bearing arrangement
As a rule, two bearings arranged at a specific distance from each other are required to guide and support a rotating machine part. There are two important variations of bearing arrangement: The locating/floating bearing arrangement and the preloaded bearing arrangement.

Locating/floating bearing arrangement
Two radial bearings are seated on the shaft and in the housing (Figure 2). The distance between the two bearing points is determined by manufactured conversion parts within the framework of the tolerance. In addition, the shaft, when heated to different temperatures or when made from different materials, does not expand in the same way as the housing. These differences must be compensated for in the bearing points. For this reason, one bearing is to be axially secured as a locating bearing on the shaft and in the housing and one bearing is to be movable in the axial direction as a floating bearing. Typical applications of the locating/floating bearing arrangement can be found in alternators and steering motors.

Single-row deep-groove ball bearings are frequently used as locating bearings. Cylindrical rolling bearings, needle bearings, and deep-groove ball bearings are usually used as floating bearings.

Double-row angular-contact ball bearings and tapered rolling bearings are also used, e.g. as wheel bearings, in the case of high radial and axial loads.

Figure 2: Locating/floating bearing arrangement

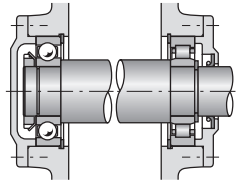
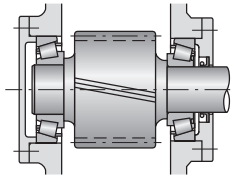


Figure 3: Preloaded bearing arrangement

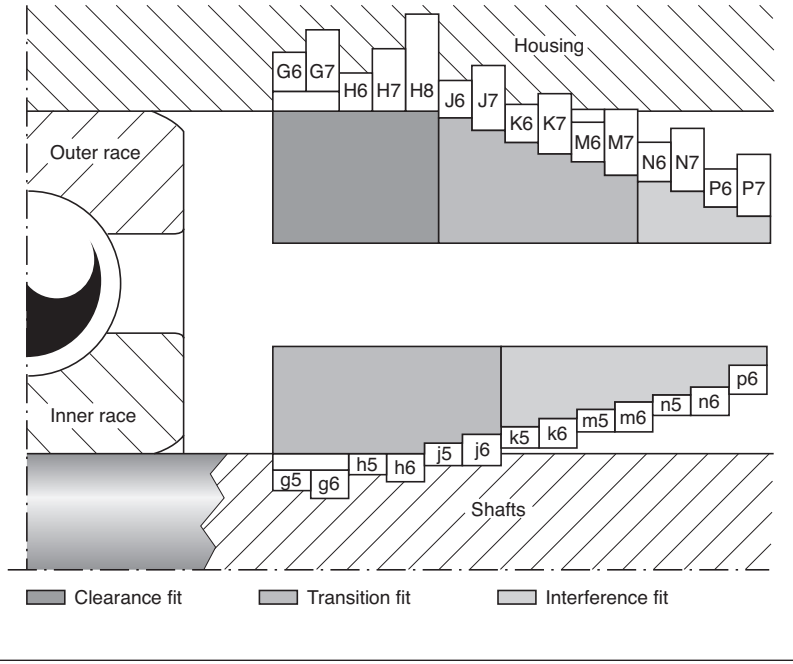


Preloaded bearing arrangement
A preloaded bearing arrangement is predominantly formed from two angular-contact ball bearings or tapered rolling bearings arranged in mirror-image fashion (Figure 3). During assembly, a bearing race is displaced on its seat surface until the bearing arrangement has attained the desired or required play or preload. Because of the possibility of play regulation, a preloaded bearing arrangement is particularly suitable for applications with close guidance, e.g. bearings in transmissions.

Tolerances and fit of bearing points
Rolling bearings essentially have negative tolerances for hole diameter, outside diameter and width, i.e. the nominal size is always the maximum limit of size.

Mounting the races on the bearing points (shaft and housing bore) is important for the installation of rolling bearings. The rolling bearings must not twist, above all tangentially, under load on the counterparts. The safest and easiest way of ensuring mounting is to make the correct choice of fit and tolerances so that the load capacity of the bearing can be fully utilized. Depending on the tolerance band of the bearing seat, the fit is referred to as clearance fit, transition fit, or interference fit (Figure 4).

Figure 4: Fits of bearing seat



The load conditions of the bearing races are of great importance to the choice of fit. A distinction is made between two loads, based on the load direction and the rotation of the bearing races.

– Rotating load: The race rotates relative to the load direction, and must be firmly seated.
– Concentrated load: The race is stationary relative to the load direction, and can have a close clearance fit to interference fit with its counterpart.

Because of the small thicknesses of the bearing races, the form variations of the seats are transferred to the raceways. The counterparts should therefore demonstrate as much form quality as possible, such as concentricity, cylindricity, and runout.

Calculation of load capacity

It is necessary to distinguish between static and dynamic load capacity when calculating the load capacity of a rolling bearing. The basic principles for calculating the static and dynamic load capacities are set out in ISO 76 [10] and ISO 281 [11].

Static load capacity
If a rolling bearing is subjected to load while stationary or while rotating at low speed, i.e. $n\,d_m \le 4{,}000\,\text{mm} \cdot \text{min}^{-1}$ (n rotational speed, d_m mean value of bore and outside diameters), it is considered under static load.

If a bearing is subjected to load in both the radial and the axial directions, the static equivalent bearing load P_0 is formed from this:

$$P_0 = X_0\,F_r + Y_0\,F_a \text{ in N}$$

where
X_0 Radial factor; X_0 = 0.6 for single-row deep-groove ball bearing.
Y_0 Axial factor; Y_0 = 0.5 for single-row deep-groove ball bearing.
F_r Radial load in N.
F_a Axial load in N.

At $P_0 < F_r$, $P_0 = F_r$ must be reckoned on. For other bearing types, refer to the specifications in the catalogs of rolling-bearing manufacturers for the factors X_0 and Y_0.

As the measure of the static load capacity, a ratio

$$f_S = \frac{C_0}{P_0}$$

is formed, where C_0 is termed the static load rating. C_0 is the load at which permanent total plastic deformation of rolling elements and raceways at the contact point subject to greatest load amounts to 0.0001 of the rolling-element diameter. Catalogs specify C_0 for all rolling bearings. The calculation of the static load rating C_0 is shown in the standard ISO 76.

In the case of normal requirements, the characteristic figure $f_s = 1$ can be approved. The requirement of lower deformation (< 0.0001) corresponds to a greater characteristic figure $f_s > 1$.

Dynamic load capacity
Calculating the dynamic load capacity is based on the method of the standard ISO 281. This refers to the service life of a rotating rolling bearing under load, where the running surfaces can be subject to material fatigue. The important characteristic figure of load capacity is the dynamic load rating C. This specifies the load of the rolling bearing for a nominal bearing service life of one million revolutions.

For the purpose of calculating the nominal service life, the following applies in accordance with ISO 281:

$$L_{10} = 10^6 \left(\frac{C}{P}\right)^p \quad \text{in revolutions,}$$

$$L_{10h} = \frac{10^6}{60n}\left(\frac{C}{P}\right)^p \quad \text{in hours.}$$

L_{10} Nominal service life achieved or exceeded by 90 % of a larger batch of identical bearings.
C Dynamic load rating in N – specified in rolling-bearing catalogs.
P Dynamic equivalent bearing load in N.
p Exponent, p = 3 for ball bearings, p = 10/3 for rolling bearings.
n Rotational speed in min^{-1}.

The dynamic equivalent load P is defined as the imaginary load constant in size and direction which influences the service life, such as the actually acting radial load and axial load. It is calculated from:

$$P = X\,F_r + Y\,F_a \text{ in N}$$

where
F_r Radial load,
F_a Axial load,
X Radial factor,
Y Axial factor.

The radial factor X and axial factor Y are dependent on bearing type, size, bearing clearance, and load ratio, and are specified in ISO 281 and in rolling-bearing catalogs.

Modified service life

In addition to the nominal service life, an extended modified service life L_{na} has been introduced in the standard ISO 281, for which operating conditions can also be included in the calculation:

$$L_{na} = a_1 \, a_2 \, a_3 \, L_{10} ,$$

where

a_1 Coefficient for survival probability, e.g.:
90 %: $a_1 = 1$;
95 %: $a_1 = 0.62$.

a_2 Coefficient for special bearing design such as internal construction and materials.

a_3 Coefficient for operating conditions, such as bearing lubrication and operating temperature.

The coefficients a_2 and a_3 are not mutually independent, and are frequently combined as coefficient a_{23}:

$$L_{na} = a_1 \, a_{23} \, L_{10} .$$

Numerous systematic examinations and experiences from practice make it possible to quantify the influences of bearing materials and operating conditions on the achievable service life of rolling bearings. Diagrams and computing programs are provided by rolling-bearing manufacturers to calculate the coefficients a_2, a_3 and a_{23}.

References

[1] DIN 623: Rolling bearings – Fundamental principles:
Part 1: Designation, marking.
Part 2: Graphical representation of rolling bearings.
[2] ISO 492: Rolling bearings – Radial bearings – Dimensional and geometrical tolerances.
[3] DIN 620-1: Rolling bearings; gauging methods for dimensional and running tolerances.
[4] DIN 620-2: Rolling bearings; tolerances for rolling bearings; tolerances for radial bearings.
[5] DIN 620-3: Rolling bearings; tolerances for thrust bearings.
[6] DIN 620-4: Rolling bearings – Rolling bearing tolerances – Part 4: Radial internal clearance.
[7] DIN 620-6: Rolling bearings – Rolling bearing tolerances – Part 6: Chamfer dimension limits.
[8] DIN EN ISO 683-17: Heat-treated steels, alloyed steels and free-cutting steels – Part 17: Ball and roller bearing steels.
[9] ASTM A295: High-Carbon Anti-Friction Bearing Steel.
[10] ISO 76: Rolling bearings – Static load ratings.
[11] ISO 281: Rolling bearings – Dynamic load ratings and rating life.

Seals

Seal technology

Function

The function of a seal is to separate two different media from each other. For economic reasons it is necessary to seal existing sealing points with simple geometries. The O-ring assumes the simplest of these geometries.

Classification

A distinction is made between static and dynamic seals (Figure 1). In the case of a static seal there is no relative movement between seal and machine parts; in the case of a dynamic seal there is relative movement between seal and machine parts. This can exist both between seal and shaft and between seal and housing.

100 % tightness can be achieved with static seals, but this for the most part cannot be achieved with dynamic seals. A “leakage rate”, which may not be exceeded, is defined for dynamic seals. When a seal complies with this rate, it is said to be subject to technical tightness.

Materials for seals

In the main elastomers but also other materials are used for seals in automotive applications. Elastomers are used in a diverse range of applications – as tires, in door seals, as axle boots, in fuel hoses, or as assorted seals in the engine compartment.

Requirements

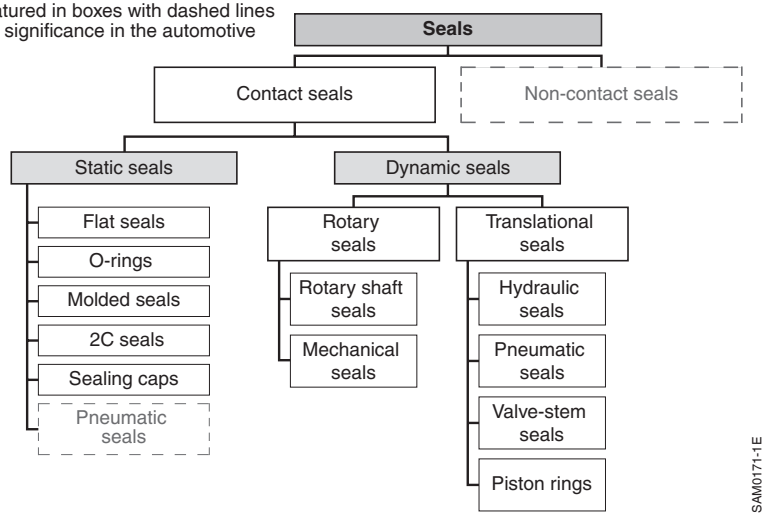
Elastomers, like all organic-chemicals, are not subject to unlimited use. External influences, e.g., different media, oxygen or ozone as well as pressure and temperature, alter the material properties and hence their behavior. Elastomers can swell, shrink, harden, crack, or even break.

The complete sealing system must always be taken into consideration in the design of a new seal. This comprises the following seven points:

– seal geometry,
– system pressure (averaged pressure and short-time peak pressures),

Figure 1: Types of seals

Seals featured in boxes with dashed lines are of no significance in the automotive sector.



– system temperature (averaged temperature and short-time peak temperatures),
– gap to be sealed,
– roughness of the mating face,
– medium to be sealed,
– relative speed between seal and machine part.

Applications
The matching sealing system is designed in line with the prevailing requirements. An overview of the existing solutions and where they are used is provided in Figure 2.

Figure 2: Seal applications in a motor vehicle

1 Air-conditioning system
- O-rings
- PTFE seals
- Rotary seals

2 Heat management
- Special rotary seals
- Multi-component seals
- Sealing rings
- O-rings
- PTFE components

3 Cylinder head
- Metal gaskets
- Multi-layer seals

4 Fuel-injection system
- O-rings
- Customer-specific sealing rings
- PTFE molded rings

5 Exhaust-gas recirculation
- EGR seal

6 Battery
- Elastomer pressure-limiting valve

7 Safety-relevant components (airbag)
- Customized two-component parts

8 Electrics
- Customized housing cover made of thermoplastic material or in combination with silicone

9 Electronic control units
- Specially developed multi-layer seals

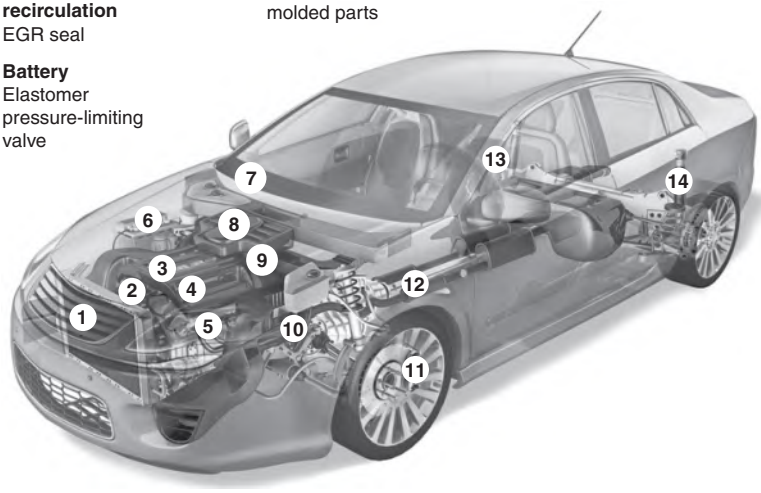
10 Engine, transmission, steering, among others
- sealing caps

11 Brake system
- O-rings
- Customer-specific molded parts

12 Drivetrain and transmission
- Rotary seals
- Guide rings
- Back-up rings
- O-rings
- Sealing washers
- Sealing caps

13 Windows and doors
- Extruded elastomer sections

14 Chassis systems
- Sealing-edge ring
- O-rings
- Buffer ring
- Back-up rings
- Vane seal



SAM0172E

O-ring

With the O-ring, design engineers are provided with an efficient and economic sealing element for a huge variety of different applications. It is used primarily as a static sealing element, but can also be used in dynamic applications. Cost-effective manufacturing processes and easy handling make the O-ring the most commonly used seal.

A large selection of elastomer materials for standard and special applications enables virtually all liquid and gaseous media to be sealed.

Description

O-rings are continuously vulcanized in molds. They are characterized by their ring shape with a circular cross-section. The O-ring is defined in its dimensions by the inside diameter d_1 and the cross-section d_2 (Figure 1).

Thanks to its simple geometry the O-ring has many benefits:
- symmetrical cross-section,
- simple, compact design,
- self- and double-acting,
- simple calculation and definition of the groove,
- continuous groove design,
- large choice of materials,
- wide range of applications.

Applications

O-rings are used as primary sealing elements, as tensioning elements for rubber-energized hydraulic seals, and as wipers (secondary sealing element). They therefore cover a huge range of applications.

Whether it be as an individual seal for a repair or as a quality-assured sealing element in automobile manufacturing or mechanical engineering – there is today no field in industry in which the O-ring is not used. The O-ring is used mainly in static seals
- as radial-static sealing, e.g., in bushings, covers, pipes, and cylinders (Figure 2),
- as axial-static sealing, e.g., in flanges, plates, and plugs (Figure 3).

Dynamic use is only possible at low load levels. It is limited by the speed and the pressure to be sealed
- to sealing e.g. back- and forward-moving pistons, rods, and plungers,
- to sealing e.g. slowly slewing, rotating or screw-like movements on shafts, spindles, and rotary transmission leadthroughs.

Operating concept

O-rings are self- and double-acting sealing elements. The pressure forces caused by installation in the radial or axial direction produce the initial tightness. They are

Figure 1: O-ring dimensions
d_1 Inside diameter,
d_2 Cross-section.

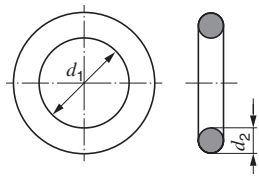


Figure 2: Radial installation of an O-ring
d_1 Inside diameter,
d_2 Cross-section,
d_3 Groove-base diameter,
d_6 Installation-space outside diameter.

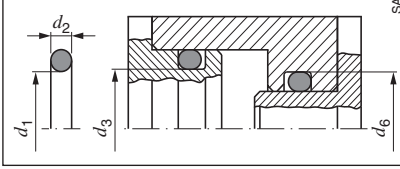
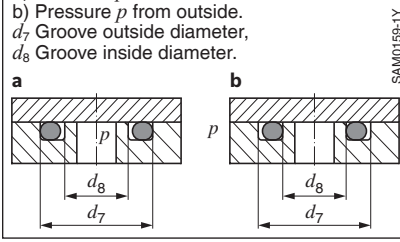
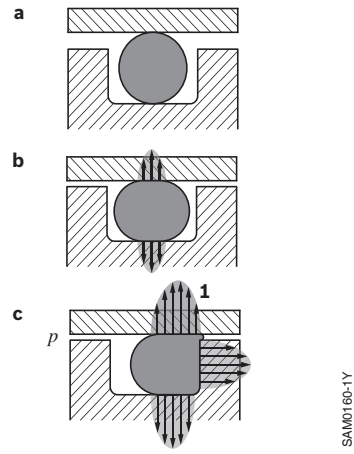


Figure 3: Axial installation of an O-ring
a) Pressure p from inside,
b) Pressure p from outside.
d_7 Groove outside diameter,
d_8 Groove inside diameter.



superimposed by the system pressure. This creates a total sealing pressure, which increases as the system pressure increases (Figure 4). The O-ring behaves under pressure similarly to a liquid with high surface tension. In this way, the pressure is uniformly transmitted to all sides.

Figure 4: Pressure forces on the O-ring
a) Without pre-pression,
b) Pressure forces with pre-pression,
c) Pressure forces with pre-pression, with system pressure.
1 Pressure path in the O-ring.
p System pressure.



Materials

Equipment manufacturers and operators expect sealing systems to function without leaks and to have long service lives. In the interests therefore of finding the ideal seal solution in an individual application, the choice of materials is, as well as the correct design, of crucial importance. The specific material properties will be discussed elsewhere (see Elastomers), since this apply equally to all seals.

Sealing gap

Excessively large sealing gaps run the risk of gap extrusion (Figure 5), which can cause the O-ring to be destroyed.

The permissible radial sealing gap S between the parts to be sealed is dependent on the system pressure, the cross-section, the media temperature, and the Shore hardness of the O-ring. Table 1 sets out recommendations for the permissible gap dimension S as a

Figure 5: Sealing gap
p System pressure,
S Sealing gap.

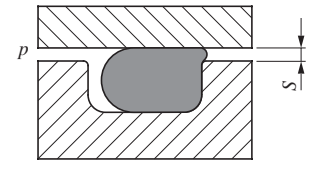


Table 1: Gap dimensions

O-ring cross-section d_2 [mm]	up to 2	2…3	3…5	5…7	>7
O-rings with hardness 70 Shore A					
Pressure p [MPa]			Gap S [mm]		
≤3.50	0.08	0.09	0.10	0.13	0.15
≤7.00	0.05	0.07	0.08	0.09	0.10
≤10.50	0.03	0.04	0.05	0.07	0.08
O-rings with hardness 90 Shore A					
Pressure p [MPa]			Gap S [mm]		
≤3.50	0.13	0.15	0.20	0.23	0.25
≤7.00	0.10	0.13	0.15	0.18	0.20
≤10.50	0.07	0.09	0.10	0.13	0.15
≤14.00	0.05	0.07	0.08	0.09	0.10
≤17.50	0.04	0.05	0.07	0.08	0.09
≤21.00	0.03	0.04	0.05	0.07	0.08
≤35.00	0.02	0.03	0.03	0.04	0.04

function of the O-ring cross-section and the Shore hardness. The table applies to elastomer materials, except polyurethane and FEP- or PFA-sheathed O-rings. In the case of pressures greater than 5 MPa for inside diameters larger than 50 mm and pressures greater than 10 MPa for inside diameters smaller than 50 mm, it is necessary to provide back-up rings (Figure 6).

Figure 6: Installation of an O-ring
a) Installation without back-up ring,
b) Installation with back-up ring.
p System pressure,
S Sealing gap.

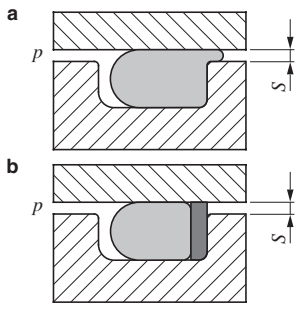
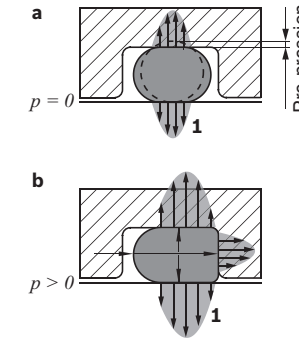


Figure 7: Sealing-pressure path in the O-ring
a) System pressure $p = 0$,
b) System pressure $p > 0$.
1 Sealing-pressure path.



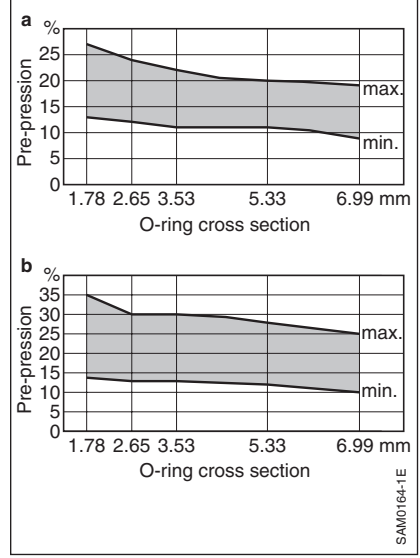
Groove filling
To avoid damaging effects on the sealing function, it is important to take into consideration the groove filling for an installed O-ring. It should in the installed state where possible not exceed 85 % so as to accommodate any thermal expansion of the O-ring, volume swelling due to media contract, and influences of tolerances.

Pre-pression
The pre-pression (Figure 7) serves among other things
– to achieve the initial tightness,
– to bridge production-related tolerances,
– to ensure defined frictional forces,
– to compensate the compression set,
– to provide compensation for wear.

Depending on the application, the following values, referred to the cross-section (d_2), are recommended for the pre-pression: for dynamic use 6 to 20 %, for static use 15 to 30 %.

For the design of grooves the guide values for pre-pression can be taken from the

Figure 8: Permissible range of pre-pression as a function of cross-section
a) Hydraulic, dynamic load,
b) Hydraulic and pneumatic, static load.



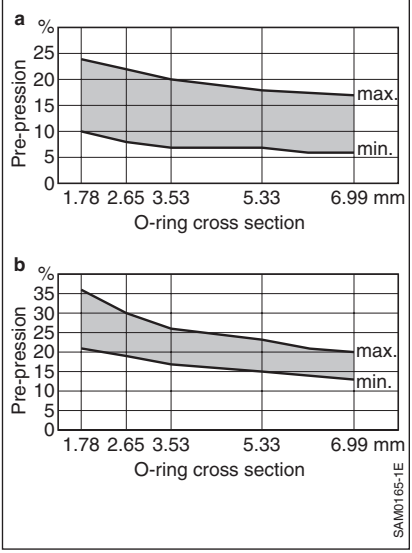
diagrams in Figure 8 and Figure 9. These taken into account in accordance with ISO 3601-2 [1] the dependence on loads and the cross-section.

Surfaces

Elastomers adapt to irregular surfaces under pressure. For gas- or liquid-tight connections, however, minimum demands must be made of the surface quality of the surfaces to be sealed. Basically, for instance, scores, scratches, shrink holes, and concentric or spiral machining marks are not permitted.

Dynamic mating faces are subject in terms of surface quality to more stringent demands than static seals. There are as yet no standardized definitions for describing mating faces. The specification of the R_a value (roughness average) is in practice not sufficient to assess the surface quality. Manufacturer recommendations therefore contain different terms and definitions among others in accordance with DIN 4768 [2] and DIN EN ISO 4287 [3].

Figure 9: Permissible range of pre-pression as a function of cross-section
a) Pneumatic, dynamic load,
b) Axial-static load.



General technical data

O-rings can be used in a wide range of applications. Temperature, pressure and media determine the choice of suitable materials. To be able to assess the suitability of the O-ring as a sealing element for a given application, it is necessary to consider the interaction of all the operating parameters. Further information can also be obtained in the calculation program for O-rings of Trelleborg Sealing Solutions [4].

Usage criteria

Operating pressure

Static usage

In static usage the following values apply to the operating pressure:

– operating pressure < 5 MPa for inside diameters > 50 mm without back-up ring.
– operating pressure < 10 MPa for inside diameters < 50 mm without back-up ring (depending on material, cross-section and gap dimension).
– operating pressure < 40 MPa with back-up ring.
– operating pressure < 250 MPa with special back-up ring.

The permissible gap dimensions must be observed here.

Dynamic usage

In dynamic usage the following values apply to the operating pressure:

– operating pressure < 5 MPa in back- and forward-moving usage, without back-up ring.
– higher pressures with back-up ring.

Speed

In back- and forward-moving usage speeds of up to 0.5 m/s are permitted, in the case of rotating seals also up to 0.5 m/s. This applies in each case depending on material and application.

Temperature

Depending on material and media resistance, usage in the temperature range of –60 to +325 °C is permissible.

The short-time peak and continued-use temperatures and the operating time must be taken into account when assessing the

usage criteria. In the case of rotating usage the temperature increases caused by frictional heat must be observed.

Media
Virtually all liquids, gases and chemicals can be sealed with a large variety of materials with different properties. Details on the choice of suitable materials can be found elsewhere (see Elastomers). Further information can be found on the Internet, e.g, in the Chemical Compatibility database [5].

Notes on installation
Before starting installation, it is necessary to check the following points:
– Lead-in chamfers fashioned as per drawing?
– Internal bores deburred and rounded?
– Machining residues (e.g., swarf, dirt, foreign particles) removed?
– Thread crests covered?
– Seals and components greased or oiled? Media compatibility with elastomer must be ensured. It is recommended to use the medium to be sealed for lubrication. No lubricants with solid additives (e.g., molybdenum disulfide, zinc sulfide) may be used.

An installation-oriented design can eliminate potential causes of seal failure (examples in Figures 10, 11 and 12). Since O-rings are always installed with interference, lead-in chamfers and edge roundings must be provided.

Installation by hand
– Do not use any sharp objects.
– Look out for twisting, use aids and appliances for correct positioning.
– Use fitting aids wherever possible.
– Do not overstretch O-rings.
– Do not stretch O-rings made of extruded round cord over the joint.

Figure 10: Rod installation with O-ring

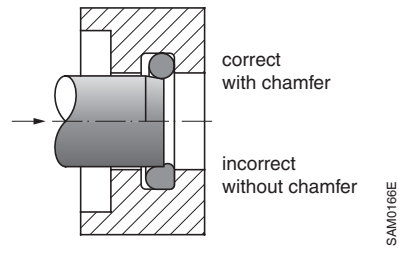


Figure 11: Piston installation with O-ring

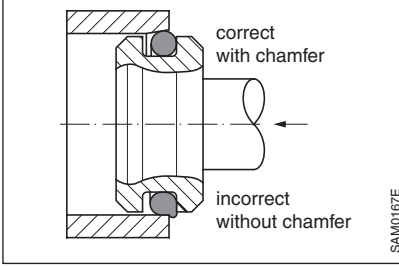
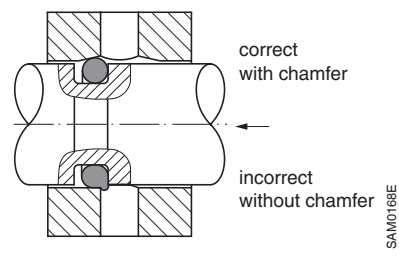


Figure 12: O-ring installation over cross-bores



Installation over threads, shafts, and similar
If the O-ring has to be guided during installation over threads, shafts, keyways or similar, a mounting sleeve will be required. This should not have any sharp edges or burrs and can be made of soft metal or plastic.

Lead-in chamfers
The minimum lengths of the lead-in chamfers are specified in Table 2 as a function of the cross-section d_2 (see also Figure 13 and Figure 14). The surface roughness of the lead-in chamfer is specified as:
$R_z \leq 6.3$ µm; $R_a \leq 0.8$ µm.

Installation types and information on installation-space design
Installation types
O-rings can be used in a variety of applications in components. The subsequent fitting situation must already be taken into account in the design. Sharp edges and bores should be circumvented so as to avoid damage during fitting. In the case of long sliding movements, the seal seat must if possible be stepped or the O-rings arranged in such a way that they only cover short mounting distances. Otherwise there is the risk of twisting.

Radial installation (static and dynamic)
The O-ring size in the case of rod seals (internally sealing) must be chosen in such a way that the O-ring outside diameter ($d_1+2\,d_2$) is at least equal to or greater than the installation-space outside diameter d_6 (Figure 2).

The O-ring size in the case of piston seals (externally sealing) must be chosen in such a way that the inside diameter d_1 is equal to or less than the groove-base diameter d_3 (Figure 2).

Axial installation (static)
In the case of axial-static installation, the pressure direction must be taken into account in the choice of the O-ring size (Figure 3). For internal pressure the O-ring outside diameter should chosen as being equal to or greater than the groove outside diameter d_7. For external pressure the O-ring inside diameter should be chosen as being less than the groove inside diameter d_8.

Elongation and upsetting
Radial installation of piston seal
When the O-ring is used as a piston seal (externally sealing), the nominal inside diameter d_1 of the O-ring (see Figure 2) should be elongated for dynamic applica-

Table 2: Lead-in chamfers

Lead-in chamfers Length Z min. [mm]		O-ring cross-section d_2 [mm]
15 °	20 °	Standard dimensions converted from inches (metric standard dimensions)
2.5	1.5	up to 1.78 (1.80)
3.0	2.0	up to 2.62 (2.65)
3.5	2.5	up to 3.53 (3.55)
4.5	3.5	up to 5.33 (5.30)
5.0	4.0	up to 7.00
6.0	4.5	over 7.00

Figure 13: Lead-in chamfer for bores and pipes
Z Lead-in chamfer.

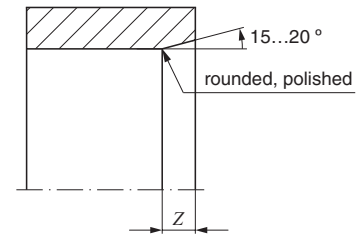
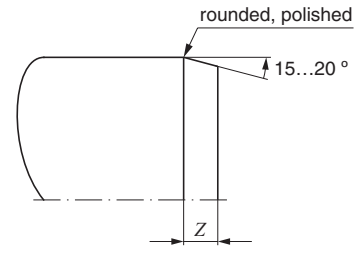


Figure 14: Lead-in chamfer for shafts and rods
Z Lead-in chamfer.



tions between 2 % and 5 % and for static applications between 2 % and 8 %. In the case of O-rings with an inside diameter d_1 < 20 mm it is not always possible to maintain this, which can result in a larger elongation range. To minimize the elongation range and the maximum stretching, it is necessary to minimize the diameter in the groove base d_3 (see Figure 2) and to impose less stringent demands on minimum stretching.

In dynamic applications it is important not to exceed a maximum elongation of 5 % so as to avoid detrimental effects on the sealing function. Basically, exceeding these recommended values will lead to a greater reduction of the O-ring cross-section, which as a result may impact on the service life of the O-ring.

Radial installation of rod seal
If the O-ring is used as a rod seal (internally sealing), the outside diameter of the O-ring (d_1+2d_2) should be at least equal to or greater than the outside diameter of the installation space (groove base) d_6 (see Figure 2) so as to achieve upsetting of the O-ring outside diameter. The outside diameter of the O-ring should for O-rings with a diameter d_1 > 250 mm not exceed 3 % of the groove outside diameter and 5 % for O-rings with a diameter d_1 < 250 mm. For O-rings with a diameter d_1 < 20 mm, this is not always possible on account of the tolerance position, which may result in greater upsetting of the outside diameter. Basically, exceeding these recommended values will lead to a marked increase in the O-ring cross-section, which as a result may impact on the service life of the O-ring.

Axial-static installation
If the O-ring is used as an axial-static seal, the pressure direction should be taken into account in the choice of the O-ring size (Figure 3). For O-ring pressurization the O-ring size should be chosen in such a way that prior to pressurization the O-ring fits closely on the groove flank on the non-pressure side. For internal pressure the O-ring should be chosen in such a way that the outside diameter (d_1+2d_2) of the O-ring is equal to or slightly greater (max. approximately 1 to 2 %) than the groove outside diameter d_7. For external pressure the O-ring should be approximately 1 to 3 % smaller than the groove inside diameter d_8.

Back-up rings
Back-up rings have no sealing functions, but, as the name suggests, are protective and supporting elements made of extrusion-resistant materials with a predominantly rectangular cross-section. They are installed together with an elastomer seal, as a rule with an O-ring, into a groove for static application. The close fit between back-up ring and bore or rod prevents the O-ring, which is under pressure, from extruding into the sealing gap.

Advantages
– Use of O-rings in high-pressure applications,
– use of O-rings of low hardness,
– compensation of large radial gap dimensions,
– externally and internally sealing application possible,
– used for static and back- and forward-moving or slowly rotating movements,
– compensation of gap enlargement by thermal expansion,
– static and dynamic applications.

Molded seal

Application

Downsizing and weight reduction in vehicles are imposing increasingly more stringent demands on the sealing systems. Engine components and add-on parts are increasingly being manufactured from plastics. Owing to the lower rigidity levels, plastic components are subject to greater deformation than metal components (e.g., axial and radial tolerances, tolerances due to high pressures).

Diecast metal components are as a rule not remachined, entailing greater component tolerances accordingly. These demanding requirements can be satisfied very well with elastomer sealing systems (Figure 1).

Sealing system

Influencing variables

Figure 2 shows the significant influencing variables for a sealing system. It depicts a molded seal for sealing between housing and cover. The system's requirements profile can be divided into the following stress types: mechanical (pressure, vibrations), chemical or physical (fluids or media), thermal, and electrical.

Design features

When it comes to the design features, the installation space (Figure 2, A1) is considered with its tolerances, flatness, the surface texture, the sealing gap, and the radius and flank angles. The installation space consists of the seals partners groove and counter-sealing face. The profile of the molded seal (Figure 2, A2) is matched under the criteria of compression and the clearance examination to the installation space.

Figure 1: Design examples
a) I-profile,
b) Double-I-profile,
c) T-profile.

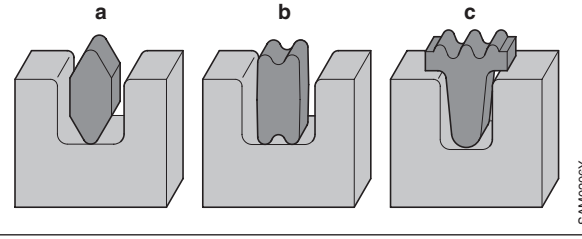


Figure 2: Significant influencing variables for a sealing system
1 Molded seal (sealing between housing and cover).
A1 Installation space,
A2 Profile of molded seal.

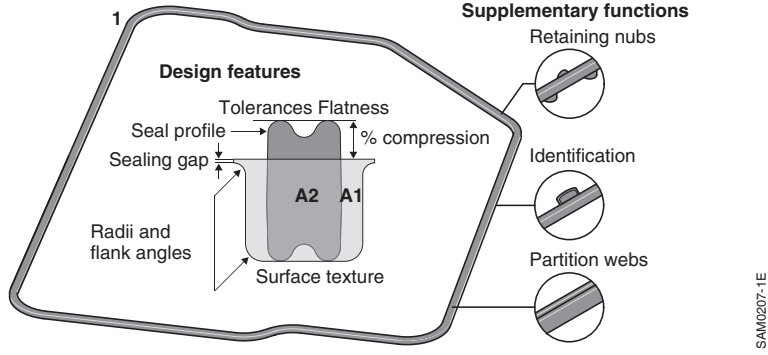
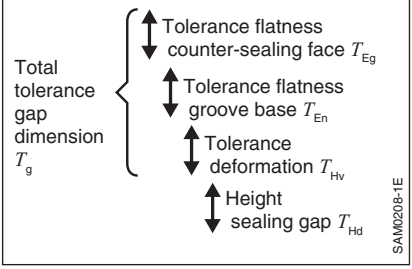


Figure 3: Relationship between gap dimension and sealing gap

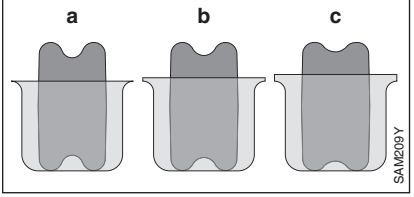


Additional functions can also be used on the molded part. Consequently, retaining nubs are used to provide transport retention for example during fitting, an identification for traceability, and partition webs for locating a leak.

The overview in Figure 3 shows the relationship between gap dimension and sealing gap. The gap dimension T_g is the sum of the component tolerances consisting of flatness of groove base T_{En} and counter-sealing face T_{Eg}, the bushing protrusion T_B and the entire widening and bending-up T_A (T_A and T_B are summarized in Figure 3 as Tolerance Deformation T_{Hv}). It is a theoretical dimension and is divided into three cases (Figure 4). If a sealing gap H_d is present in the design, this must be taken into account in the configuration of the compression.

Figure 4: Overview: minimum, nominal and maximum clearance

a) Maximum sealing, minimum installation, minimum clearance.
b) Nominal sealing, nominal installation, nominal clearance.
c) Minimum sealing, maximum installation, maximum clearance.



Finite-element method

The finite-element method is a numeric analysis method for obtaining estimates for complex technical problems (see Finite-element method). All FEM calculations are based on idealized input data. A calculation is expedient only if the input variables for the material model, the system's requirements profile (e.g., pressure, temperature, installation) and the geometry are known precisely. For example, the overall elongation of a seal can be determined after installation at room temperature and a defined overpressure. The results are shown in graphic form with the areas with differing elongation being colorized.

Flat seals

Application
Flat seals (gaskets) and their limits of use are very much dependent on the material used. Here, the applications vary in vehicle manufacturing from hot fluids, through strong chemical stress, to high pressures. Depending on the requirement, the flat seal can be sheathed in metal or even made up entirely of metal.

The different types of flat seal are as follows:
– Less hard-wearing: Paper, compressed fiber, elastomer, hard- and elastomer-bonded cork.
– Medium hard-wearing: Bonded glass-fiber, aramide-fiber, carbon-fiber and mineral-fiber seals (soft-material seals), elastomer-metal composite.
– Strongly hard-wearing: Expanded graphite, metal.
– Corrosion-resistant: Solid, expanded, filled PTFE; corrosion-resistant metal, elastomer-metal composite.

Variants
As the above list shows, there is a large variety of flat seals. Only the most important of this will be discussed in the following.

Cylinder-head gaskets
Cylinder-head gaskets (Figure 1) are fundamental static seals in the engine compartment. The complex flat seal as a rule consists of one or more coated metal layers with the possibility of having elastomer areas vulcanized on. The cylinder-head gasket not only seals against combustion gas, oil and coolant, but also acts as a power-transmission element between the cylinder head and the crankcase.

Soft-material seals
Soft-material seals made of elastomer-fiber composite materials are, in terms of their basic principle, some of the oldest seals. They depend first and foremost on the quality of their fibers and are very pressure- and temperature-resistant. They are used for example as intake-manifold seals and in exhaust-gas recirculation valves. Natural rubber, NBR, SBR, CR, HNBR and FKM are used (see Elastomers) as binders in elastomer-bonded soft-material seals. For higher stresses simply flat seals with expanded graphite or pure metal gaskets can be used.

Metal-bead gaskets
In the case of narrow sealing webs or low surface pressure, a good solution is to use metal-bead gaskets (used for example as cover gaskets in pressure applications). The surface pressure is reduced on account of the beads to a linear pressure. This allows a higher surface pressure with consistent screw force. An elastomer coating applied to the carrier plate also facilitates the micro-sealing effect.

Elastomer-metal flat seal
The flat seal made of an elastomer-metal composite is particularly suitable for preventing micro-leaks. This inherently sta-

Figure 1: Example of a cylinder-head gasket for a cylinder bank of a 6-cylinder engine with two exhaust and two intake valves

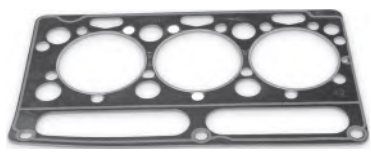
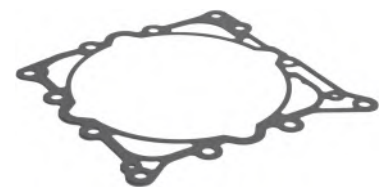
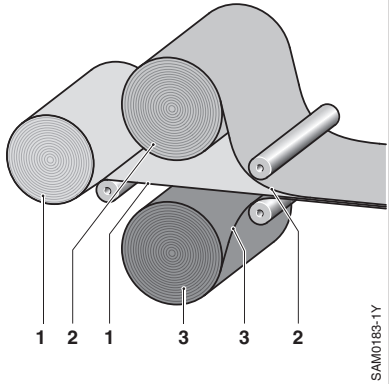


Figure 2: Flat seal in an electric-motor housing with integrated cooling



ble and at the same time adaptable seal consists of a thin metal layer with vulcanized-on elastomer layer (Figure 2). Stamped into complex geometries, the seal is ideal for use in primary frictional connection. With an elastomer-layer thickness of at least 0.2 mm, it also offers the opportunity of compensating for shrink holes and pores in the casting. The elastomer-metal flat seal is used in flange joints, pumps, automatic transmissions, and electric motors.

Figure 3: Coil production
1 Metal carrier,
2 First elastomer coil,
3 Second elastomer coil.



Sealing caps

Application
Sealing caps belong to the group of static seals. They facilitate the reliable sealing of assembly- or production-related bores at various points in the engine compartment. Design engineers can always choose between the standard solution according to DIN 443 [6] and other special types which meet the high requirements of the automotive industry.

The standard solution according to DIN 443 describes a cap made of metal which is generally pressed into the housing and provides secure sealing only by means of an additional adhesive connection. Set against these caps are sealing caps, which have an elastomer layer and do not required any additional adhesives. The elastomer layer furthermore facilitates installation without damaging the housing surface and thereby prevents micro-leak paths.

Technology
Metal sealing caps are worked in a deep-drawing process from the coil material into the finished product. The favorable elastomer-metal sealing caps can be manufactured in different ways. Possible production processes are compression molding, injection molding, lamination or vulcanization. One innovative production process involves the technology whereby multiple-layer rubber-metal connections are produced, and then stamped and deep-drawn as composite material (Figure 3).

The process of applying a binder to metal carriers and vulcanizing elastomers on metal creates secure and solid connections which can be fashioned into complex seal geometries. In this way, flat seals, seals with metal frames for automated handling, friction-optimized guide bands, and housing gaskets and sealing caps can be specifically produced.

Often, as a result of the optimized rubber-metal connection, the surfaces of the sealing faces no longer have to be machined in a complicated process. The elastomer layer improves sealing performance and, furthermore, no adhesives or sealing compounds are required in

production, enabling the cycle times in assembly to be reduced and the production process to be fashioned in a more process-reliable way. Users gain cost advantages through simpler handling, the possibility of automation in assembly, and reduced logistics expenditure. There are cost-reduction potentials, particularly in respect of injection-molded sealing caps.

Examples in practice

Engine blocks are produced by sand casting. At the end of the production process, the engine block features a number of bore holes which cannot be avoided from a production standpoint, but must nevertheless be sealed. Rubber-metal sealing caps offer an ideal solution to this since they are easily mounted, not bonded, and rustproof. The installation situation is pictured in Figure 4.

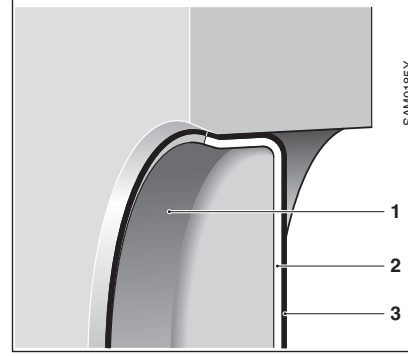
There are further possible uses in the form of static sealing of bore holes in the transmission and in steering systems. But rubber-metal sealing caps are also used in practice in the camshaft housing and crankcase and in pumps.

Variants

Rubber-metal sealing caps can be manufactured from different materials. Steel, galvanized steel and stainless steel may be used as metal materials. Possible elastomer materials are NBR, EPDM and AEM, and hence provide for a temperature range of –45 to 150 °C and different media-resistance capabilities.

Figure 4: Installation situation of sealing cap

1 Sealing cap,
2 Metal,
3 Rubber layer.



Radial shaft seal

Application

Radial shaft seals are ring-shaped sealing elements the purpose of which is to separate permanently and safely from each other oil or grease from the inside and, for example, dirt, dust or water from the outside. Radial shaft seals are used in drive systems (axles and transmissions, e.g., Simmer shaft seal), pumps, electric motors, and in machine-building.

Design

Radial shaft seals generally consist of a sealing lip of elastomer material and a strengthening ring of metal. A tension spring lends the sealing lip its constant pre-tension over the entire service life (Figure 1). The sealing lip is the end of the diaphragm (sealing cup). The sealing edge is the area of the sealing lip that touches the shaft and assumes the actual sealing function.

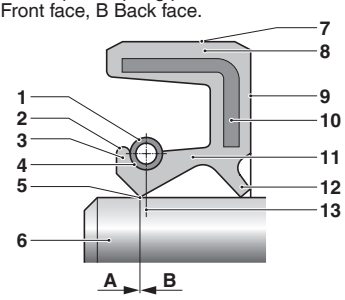
The sealing-lip geometry corresponds to today's state-of-the-art and is based on many years of application-technology experience. The sealing edge can be press-finished or manufactured at the face by mechanical cutting.

The entire radial force of the seal is created by the initial force of the elastomer sealing lip and the tensile force of the spring. The former is obtained depending on deformation from the elasticity of the material, the sealing-lip geometry, and the overlap between the shaft and the seal.

Specifically for the automotive industry the function and application-related information regarding radial shaft seals are described in DIN 3761 [7] and in ISO 6194 [8].

Figure 1: Profile section of a radial shaft seal

1 Tension spring, 2 Spring-retaining lip,
3 Sealing lip, 4 Spring groove,
5 Sealing edge, 6 Shaft, 7 Outer casing,
8 Elastomer, 9 Back casing,
10 Strengthening ring,
11 Diaphragm (sealing cup),
12 Dust lip, 13 Spring plane.
A Front face, B Back face.



Seal design

Standard types

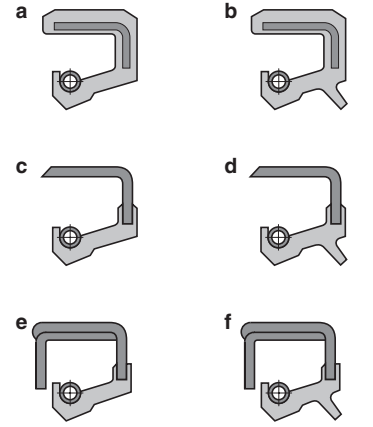
Radial shaft seals are available in standard types according to DIN 3760 [9] and DIN 3761 or ISO 6194. Figure 2 shows versions in which the strengthening ring is completely enclosed in elastomer and versions in which the strengthening ring is embedded in elastomer only in the area of the diaphragm and consequently forms the casing of the shaft seal (adhesion part to which the elastomer adheres).

Special types

Special types are used when application parameters are outside the scope of DIN 3760 and DIN 3761 or ISO 6194. This can be increased speed above 10 m/s, pressure above 0.05 MPa, grease used

Figure 2: Radial shaft seals in standard types

Standard types acc. to DIN 3760 and DIN 3761 or ISO 6194.
a) DIN 3760 Type A, b) DIN 3760 Type AS
c) DIN 3761 Type B, d) DIN 3761 Type BS
e) DIN 3761 Type C, f) DIN 3761 Type CS



instead of oil as the medium to be sealed, or increased dirt accumulation. It is recommended to use a special seal in all these situations. Figure 3 shows some special types by way of example.

Main features

The main features to be observed in any choice of seal is the outer casing, the strengthening ring, the tension spring, and the material.

Outer casing

The outer casing can be either smooth or grooved; in both cases the seal can be pressed into bores in accordance with ISO fit H8. The outside diameter of the radial shaft seal is manufactured in accordance with ISO 6194-1 (Table 1).

Metal strengthening ring

Cold-rolled sheet steel in accordance with DIN EN 10139 [10] is used for the standard version. Depending on installation conditions and environmental conditions, however, other materials such as brass and stainless steel (steel grade 1.4301 in accordance with DIN EN 10088-3 [11]) may come into play.

The primary purpose of the strengthening ring is to stiffen and reinforce the radial shaft seal. Normally, this ring may not be subjected to axial load. Should this be necessary, a special version of the strengthening ring may also be used (e.g., metal version, partly rubberized version).

Tension spring

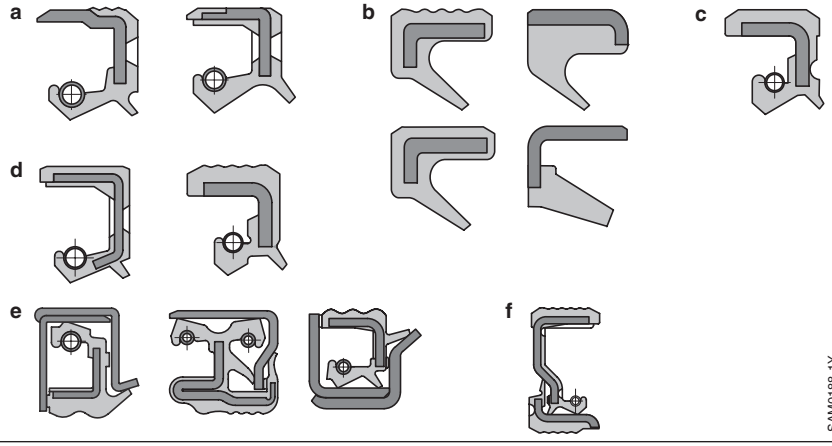
When rubber is exposed to heat, load or chemical stress, it gradually loses its original properties. This is called rubber aging. The original radial force of the sealing cup is lost. The primary purpose of the tension

Table 1: Tolerances acc. to ISO 6194-1

Outside diameter d_2 nominal [mm]	Diametrical tolerances [mm]	
	Metal housing	Rubber-coated
$d_2 <$ 50	+0.20 +0.08	+0.30 +0.15
50 < $d_2 <$ 80	+0.23 +0.09	+0.35 +0.20
80 < d_2 < 120	+0.25 +0.10	+0.35 +0.20
120 < d_2 < 180	+0.28 +0.12	+0.45 +0.25
180 < d_2 < 300	+0.35 +0.15	+0.45 +0.25
300 < d_2 < 500	+0.45 +0.20	+0.55 +0.30

Figure 3: Special types of radial shaft seals

a) Types with semi-rubberized outside diameter, b) Rotary seals without tension spring, c) Type for medium to high pressures, d) Types for medium pressures, e) Cassette seals, f) System-shaft seals.



spring therefore is to maintain the radial force.

Tests have shown that the radial force must be different depending on size range and sealing-ring type. It has also emerged from these tests that it is very important to keep the deviations of radial force during the life of the seal within tight limits. The radial force has been established by comprehensive laboratory tests.

The tension spring is coiled tightly and with pre-tension. The total force of the tension spring thus comprises partly the pre-tension and partly the force obtained from the spring rate. Using a tension spring with pre-tension offers the following advantages:
– In the event of sealing-lip wear, the part of the total radial force resulting from the spring pre-tension remains unchanged.
– Through partial softening of the spring (by heat treatment) the pre-tension can be regulated in such a way that the envisaged radial force is achieved for the respective shaft diameter.

Elastomer material of the seal

The main features for ensuring reliable functioning of radial shaft seals are both the correct design and the choice of the correct material. Therefore the choice of material must be taken into consideration as one of the most important decisions with regard to media compatibility and environmental conditions. Some material properties which are directly connected to the environmental conditions are:
– good chemical resistance,
– good resistance to heat and cold,
– good ozone and weather resistance.

Functional material requirements are among others:
– high resistance to wear,
– low friction,
– low deformation by compression,
– good elasticity.

For cost reasons, good workability is desirable as an additional property. None of the materials available today can satisfy all these requirements. The choice of material therefore always constitutes a compromise between the relative importance of the respective factors.

Table 2: Material recommendations

Materials for sealing conventional media		Material designation				
		NBR	FKM	ACM	VMQ	HNBR
		Material code				
		N	V	A	S	M
		Max. permissible continuous temperature [°C]				
Mineral lubricants	Engine oils	100	170	125	150	130
	Transmission oils	80	150	125	130	110
	Hypoid gear oils	80	150	125		110
	ATF fluids	100	170	125		130
	Pressure fluids (DIN 51524, [14])	90	150	120		130
	Greases	90				100
Fire-resistant pressure fluids (VDMA 24317, [12]) (DIN 24320, [13])	Oil-water emulsion	70			60	70
	Water-oil emulsion	70			60	70
	Aqueous solution	70				70
	Anhydrous fluids		150			
Other media	Heating fuel oils	90				100
	Water	90	100			100
	Suds	90	100			100
	Air	100	200	150	200	130

NBR Nitrile butadiene rubber, FKM Fluorine rubber, ACM Acrylate rubber, VMQ Silicone rubber. HNBR Hydrogenated nitrile butadiene rubber.

Materials and their designations are
– nitrile butadiene rubber (NBR),
– acrylate rubber (ACM),
– silicone rubber (VMQ),
– fluorine rubber (FKM),
– hydrogenated nitrile butadiene rubber (HNBR).

Hydrogenated nitrile butadiene rubber (HNBR) is a further development of conventional nitrile butadiene rubber (NBR). This material offers substantially improved heat and ozone resistance and can be used instead of acrylate rubber and in certain cases also instead of fluorine rubber. A special composition is developed for each rubber type in order to satisfy the requirements imposed on seals. In addition, still more compounds are available for some extreme conditions.

Table 2 sets out material recommendations for various applications.

Design of shaft and bore

Surface texture, hardness and machining method

The shaft design is of crucial importance both to the seal and to the service life (Figure 4). Basically, the higher the circumferential speeds the greater the hardness of the shaft should be. It is established in DIN 3760 that the shaft must exhibit a hardness of at least 45 HRC. As the circumferential speed increases, the demand for hardness rises, and at 10 m/s a hardness of 60 HRC is required.

The choice of appropriate hardness is not only dependent on the circumferential speed, but also influenced by factors such as lubrication and particles that cause wear. Poor lubrication and difficult external conditions (e.g., high dust deposits on construction machinery) therefore require a higher shaft hardness. DIN 3760 and DIN 3761 lay down maximum values for the surface roughness; a surface roughness of R_t = 1 to 4 µm (maximum roughness) is recommended. It has, on the other hand, emerged from laboratory tests that the most favorable roughness is R_t = 2 µm (roughness average R_a = 0.3 µm). Both rougher and finer surfaces cause higher friction, which results in higher temperature and increased wear. Consequently, a roughness of R_t = 2 to 3 µm (R_a = 0.3 to 0.5 µm) is to be recommended.

Friction and temperature measurements have also shown that grinding the shaft is the best machining method. Spiral grinding marks can however cause a pumping effect and leakage, which is why plunge-cut grinding should be chosen. Whole-number conditions of wheel speed to workpiece speed must be avoided here. Polishing the contact surface with emery cloth produces a surface texture which causes higher friction and temperature build-up than plunge-cut grinding.

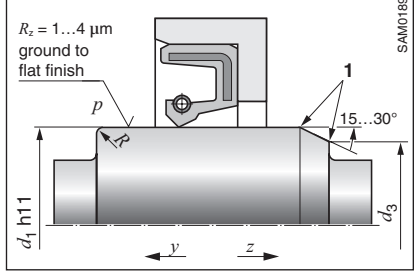
In some cases it is not possible to furnish a shaft with the hardness, surface quality and corrosion resistance required for the seal. However, this problem can be solved by fitting a separate sleeve on the shaft. In the event of wear, only the sleeve has to be replaced.

Radial run-out

Radial run-out of the shaft is to be avoided wherever possible or kept within the tightest limits. At high rotational speeds the sealing lip is at risk of not being able to follow the shaft due to its inertia. The shaft seal must be arranged in direct proximity to the bearing (e.g., roller bearing) and the bearing clearance kept as small as possible (Figure 5).

Figure 4: Shaft design

1 Edge rounded and polished, lead-in chamfer 15 to 30 °.
d_1 Diameter of shaft, tolerance h11,
d_3 Inside diameter of chamfer,
R Radius,
R_z Average peak-to-valley height,
p Pressure of medium,
y, z Direction of installation.



Deviation in concentricity
A deviation in concenticity between the shaft and the accommodating bore is to be avoided wherever possible so that the sealing lip is not subjected to load on one side (Figure 6). Depending on the direction of installation y or z, it is recommended to incorporate a chamfer or a radius. The dimensions for this are shown in Figure 4 and Table 3.

Properties of the shaft surface
The values of the moving surface are established for shaft seals in DIN 3760 and DIN 3761 or ISO 6194. The surface should exhibit the following properties:

Surface roughness
R_a = 0.2 to 0.5 µm (arithmetic roughness average),
R_z = 1 to 4 µm (average peak-to-valley height),
R_{max} (= R_t) = 6.3 µm.

Table 3: Chamfer length for the shaft end

Diameter d_1	Diameter d_3	Radius R
$d_1 <$ 10 mm	d_1 – 1.5 mm	2 mm
10 mm ≤ $d_1 \leq$ 20 mm	d_1 – 2.0 mm	2 mm
20 mm ≤ $d_1 \leq$ 30 mm	d_1 – 2.5 mm	3 mm
30 mm ≤ $d_1 \leq$ 40 mm	d_1 – 3.0 mm	3 mm
40 mm ≤ $d_1 \leq$ 50 mm	d_1 – 3.5 mm	4 mm
50 mm ≤ $d_1 \leq$ 70 mm	d_1 – 4.0 mm	4 mm
70 mm ≤ $d_1 \leq$ 95 mm	d_1 – 4.5 mm	5 mm
95 mm ≤ $d_1 \leq$ 130 mm	d_1 – 5.5 mm	6 mm
130 mm ≤ $d_1 \leq$ 240 mm	d_1 – 7.0 mm	8 mm
240 mm ≤ $d_1 \leq$ 500 mm	d_1 – 11.0 mm	12 mm

Hardness
55 HRC or 600 HV,
hardness penetration depth at least 0.3 mm.

Housing bore for the shaft seal
The housing dimensions are shown in Table 4 and Figure 7. The tolerances for the metric sizes comply with DIN 3761 or ISO 6194 so that a good press fit is achieved for a tolerance in the housing bore with ISO H8. For the imperial sizes (inches) the tolerances conform to the American standards. In installation cases where the housing bore has a different tolerance, the seal can be manufactured with a matching interference. A special fit between the seal and the bore may be required for housing bores of soft material, e.g., light alloy, in common with bearing housings with thin walls. The tolerances for seal and bore must in such cases by defined by practical tests.

If a part, e.g., a bearing, is pressed by the seal seat, this can be damaged. To avoid such damage, the seal must be chosen with a larger outside diameter than that of the bearing.

Surface roughness of the housing bore
The values for the surface roughness in the housing bore are specified in ISO 6194-1. The following values are recommended:
R_a = 1.6 to 3.2 µm,
R_z = 6.3 to 12.5 µm,
R_{max} (= R_t) = 16 to 25 µm.

Figure 5: Radial run-out

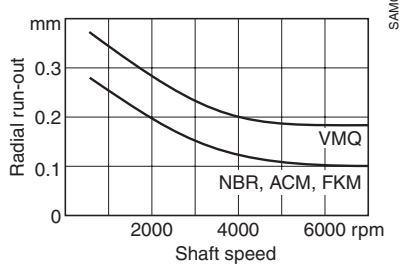
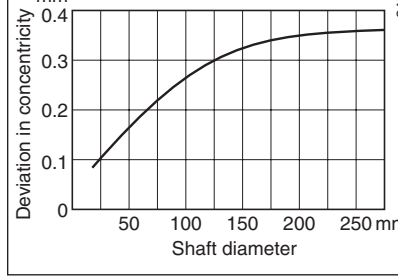


Figure 6: Deviation in concentricity



For metal-caged seals (not rubberized) or required gas tightness, a good scoring-free, flat-finish surface quality is required. If the radial shaft seal is glued in the housing, it must be ensured that no adhesive comes into contact with the sealing lip or the shaft.

Installation instructions

The following points must be observed in relation to installing radial shaft seals:

- The installation spaces must be cleaned prior to installation.
- In the case of rubber seals, shafts and seal must be greased or oiled.
- Sharp-edged transitions must be chamfered, rounded or covered.
- When pressing in, it is important to ensure that the seal is not tilted.
- The press-in force must be applied as closely as possible to the outside diameter.
- The seal must be seated centrically and perpendicularly to the shaft.
- The end face of the locating bore is usually used as the stop face; the seal can also be secured with a shoulder or a spacer.

Table 4: Housing dimensions

Width of seal b mm	b_1 min. (0.85 b) mm	b_2 min. (b + 0.3) mm	r_2 max. mm
7	5.95	7.3	0.5
8	6.80	8.3	
10	8.50	10.3	
12	10.30	12.3	0.7
15	12.75	15.3	
20	17.00	20.3	

Figure 7: Housing dimensions: installation depth and lead-in chamfer
Quantities, see Table 4.

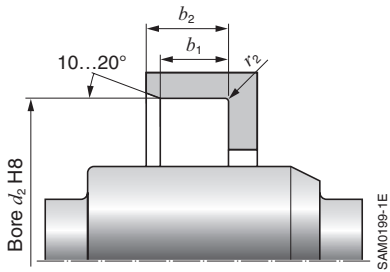


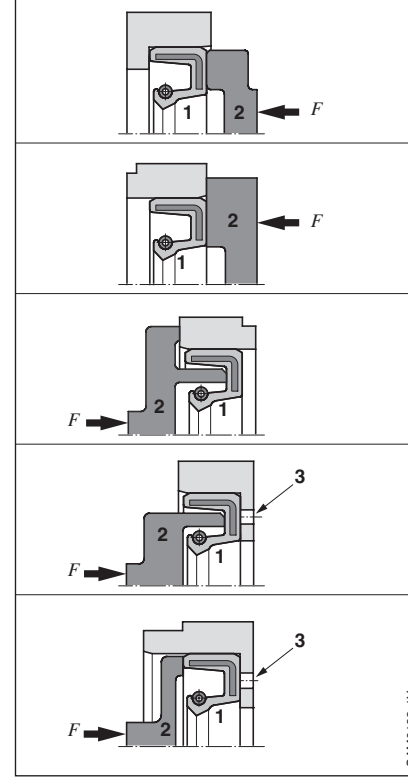
Figure 8 shows different press-in situations for the radial shaft seal with suitable mounting tools or devices.

Removal and replacement

Generally speaking, the removal of seals does not pose any difficulties. A screwdriver or similar is usually sufficient to remove them. The seal is damaged in this process. After a machine has been repaired or overhauled, new radial shaft seals should always be fitted, even if the old seals still appear to be intact.

The sealing edge of the new ring should not rest on the old running point. This can be achieved by

Figure 8: Installation aids in the mounting of radial shaft seals
1 Radial shaft seal,
2 Mounting tool,
3 Removal hole.
F Press-in force.



– replacing the shaft protective sleeve,
– pressing into the locating bore to a different depth,
– reworking the shaft and fitting a shaft protective sleeve.

Usage parameters

Overpressure

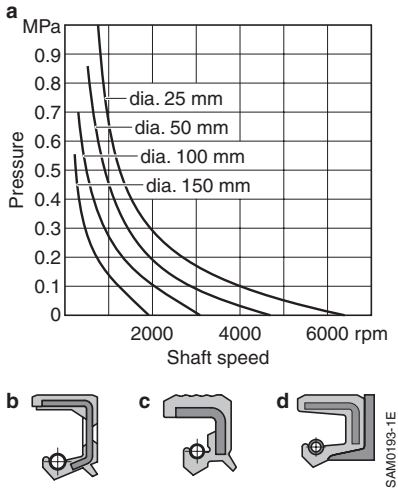
If the sealing cup is subjected to overpressure, it is pressed against the shaft, increasing the contact area of the sealing lip against the shaft. This increases friction and heat build-up. At overpressure, therefore, the guide values for the maximum permissible circumferential speed are not applicable, but instead these must be decreased in proportion to the intensity of the pressure. At high circumferential speeds, however, already overpressures of 0.01 to 0.02 MPa can cause problems. By using an additional back-up ring, it is possible to use the DIN types A, AS, C and CS pictured in Figure 2 above 0.05 MPa. The separate back-up ring is to be adapted to the back face of the sealing cup. It should not however rest on the sealing cup while there is no overpressure applied. The back-up ring must be adjusted to fit as accurately as possible.

In type TRU the strengthening ring is fashioned in such a way that it supports the sealing cup (Figure 9). Type TRP has a short and sturdy sealing lip which permits an overpressure without additional support. When a back-up ring is installed or when thrust rings are used, overpressures of 0.4 to 0.5 MPa can be permitted at moderate circumferential speeds.

At high overpressures seal types with rubber outer casings should be chosen (type TRA) so as to prevent leakage at the locating bore. At overpressure the seal is at risk of shifting in the axial direction in the housing bore (pressing out). This can be avoided by securing the seal with a shoulder, a spacer ring or a retaining ring.

Figure 9: Permissible pressure of the medium to be sealed for supported radial shaft seals and for pressure seals

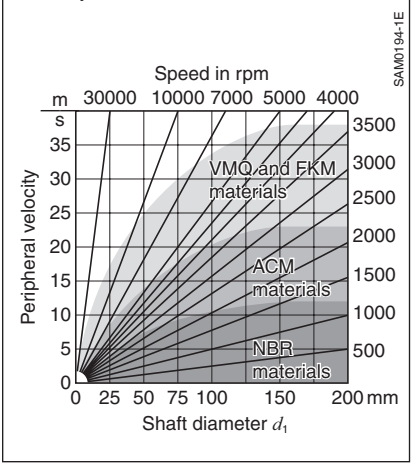
a) Diagram,
b) Type TRU,
c) Type TRP/6CC,
d) Type TRA/CB with back-up ring.



Circumferential speed and rotational speed

Different sealing-cup designs influence the level of friction and thus result in varying temperature increases. The upshot of this is that the different sealing-cup designs permit different circumferential speeds. Figure 10 features guide values for the maximum permissible circumferential speed for sealing elements without protective lip made of NBR, ACM, FKM

Figure 10: Permissible speeds in state at zero pressure acc. to DIN 3761



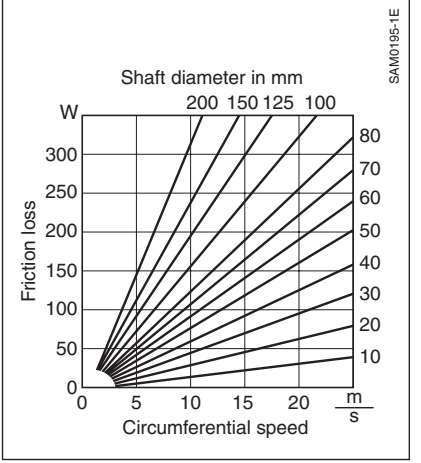
and VMQ in operation at zero pressure and where adequate lubrication or cooling of the sealing edge is guaranteed by the medium to be sealed. The permissible continuous temperatures in Table 2 must be taken into account and may not be exceeded. The bend demonstrates that larger shaft diameters permit higher circumferential speeds than smaller shaft diameters. This is because there is greater heat dissipation as the shaft cross-section increases.

Friction losses

Friction loss is often of a magnitude that must be taken into account. This applies particularly when lower power output is transmitted. Friction loss is influenced by the following factors: Seal design and seal material, spring force, rotational speed, temperature, medium, shaft design, and lubrication.

Figure 11 demonstrates which friction losses in watts are caused by a radial shaft seal without protective lip when it is installed according to the manufacturer's technical directions. In certain cases the friction loss can be reduced by special sealing-lip design. Decreasing the spring force or using a special rubber quality can also achieve the same result. In this connection it must be noted that the friction loss during the seal's "breaking-in period" is greater than that shown in Figure 11. The standard breaking-in period is a few hours.

Figure 11: Friction loss for a nitrile-rubber seal, type TRA



Pneumatic seals

Sealing against compressed air entails a variety of challenges, which are faced with pneumatic valves. Different factors from those relevant to hydraulic seals come to the fore.

As a rule, pneumatic applications are operated with an air pressure of 0.8 to 1.0 MPa, in exceptional cases up to 1.6 MPa, and for high-pressure valves sometimes even up to 5.0 MPa. Speeds are in the range of 0.2 to 1.0 m/s, and in special applications up to 5.0 m/s.

Leakage, friction and lubrication
The leakage of air plays a subordinate role. Friction has a much greater influence on the power output of a pneumatic application. Lubrication is therefore very important. Since the use of oiled air is declining, most pneumatic sealing systems are lubricated with grease on a one-off basis during installation. This lubrication must be maintained over the entire service life. One resulting requirement imposed on a pneumatic seal is hence that the lubricating grease is not wiped in the application. There are therefore differences, when compared with hydraulic seals, in the geometry of the sealing edge and the resulting pressure path.

Alternatively, seals with dry-running properties can be used. However, these are for the most part ruled out for economic reasons.

Sealing system
The most important components of a pneumatic sealing system are pictured in Figure 1.

Piston seals
Piston seals are as a rule “internal” seals and ensure that a pneumatic cylinder can move in that they prevent the medium (air or gas) from overrunning. Piston seals are seated on the moving component and are for the most part double-acting, i.e. the seal acts in both directions. It seals when the piston moves to the left and when it moves to the right. This is often manifested in a symmetrical profile.

Rod seal
Rod seals are as a rule “external” seals and ensure that no medium can escape to the atmosphere. Rod seals are seated in the housing and are for the most part single-acting. The seals acts only in one direction, with these frequently being manifested in an asymmetrical profile.

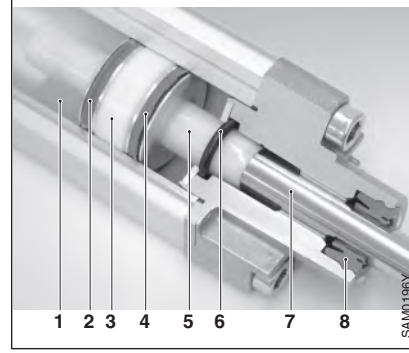
Dirt wipers
Dirt wipers protect the system against dirt, foreign particles, swarf, and moisture. This prevents premature failure of the components. Single- and double-acting wipers are used, depending on the application and sealing system.

Cushioning seal
Cushioning valves damp the impact of the piston and also assume the function of a non-return valve.

Guide band
Guide bands guide the piston and the rod. They accommodate radial forces, absorb vibrations, and prevent metal-to-metal contact.

Figure 1 shows a standard cylinder comprising the above-mentioned components.

Figure 1: Sealing system of a pneumatic cylinder with piston seal and rod seal
1 Cylinder barrel, 2 Piston seal,
3 Guide band, 4 Piston seal,
5 Piston extension, 6 O-ring, 7 Piston rod,
8 Rod seal.



Design types
Aside from classic profile seals such as the O-ring or the X-ring, there are also standard sealing elements for pneumatics. Figure 2 shows some of these seals.

Beyond simple sealing elements the trend is towards the assembly such that complete pistons are already used in many applications (Figure 3).

Materials
Different materials are used to suit the application: Elastomers, polyurethane and polytetrafluoroethylene (PTFE).

Elastomer seals
Elastomer materials (above all NBR and FKM) are, thanks to their cost effectiveness, the main material used in the production of pneumatic seals. Simply fitting is a positive side effect. Challenges for elastomer seals can be presented by high abrasion.

Figure 2: Pneumatic seal profiles for standard cylinders
a) Rod seal, single-acting,
b) Piston seal, single-acting,
c) Cushioning seal,
d) Rod-seal wiper combination
(cf. Figure 1, Profile of rod seal).

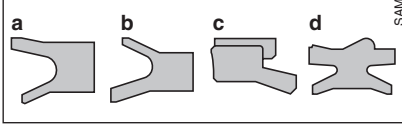
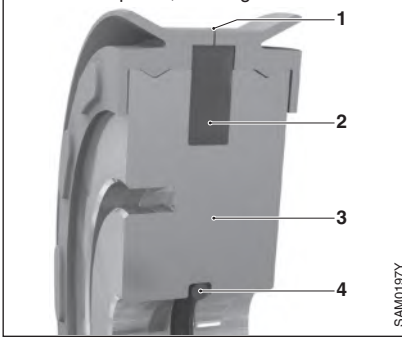


Figure 3: Pneumatic complete piston
1 Double-acting seal arrangement,
2 Magnet ring (for position sensing),
3 Aluminum piston, 4 O-ring.



Polyurethane seals
As belonging to the family of elastomers, polyurethane must be considered separately. As an abrasion-resistant and flexible material, it is often classified between elastomers such as NBR, FKM and PTFE.

PTFE seals
Compared with elastomers, PTFE is frequently not cost-effective because of the manufacturing process involved. It does however offer advantages in conditions of high pressure, wear and friction when there is no lubrication.

Applications
Pneumatic components are used in practically every branch of industry. They feature most strongly in automation engineering or process automation. Here, pneumatic units are used among others for gripping, turning, separating, suction or driving. These components ensure highly efficient, precise and economical production and assembly systems. Open and closed loop control is performed by pneumatic cylinders and valves.

Another area of application is commercial vehicles (trucks, buses, trains). Here they are used in the drive, chassis and suspension, and transmission, in among others non-return valves, pressure-control valves, and switching valves. Cylinders are used in these sectors for example for the precise control of throttle valves.

But pneumatic systems are being used with increasing regularity in other sectors such as biosciences, the food and drinks industries, the chemical process industry, and the electronics industry.

Two-component seals

Multi-component injection molding
Basically, the term multi-component injection molding covers all materials and their combinations which can be combined in the injection-molding process. The following material combinations are the most frequently encountered at present:
– thermoplastic with thermoplastic,
– thermoplastic with thermoplastic elastomer,
– thermoplastic with elastomer,
– multi-color injection molding.

Basic principles
There are various process technologies which can be used in the manufacture of multi-component molded parts. All production technologies are based on the same production process, which works with two or more injection units; however, an explicit distinction is made here between the single-stage process and the two-stage process.

Single-stage process
In the single-stage process the premolded part is transported by a handling unit from one side to the other station in the mold. The later multi-component molded part is now either ejected by means of a mold movement (e.g., ejector pin) or else removed by the handling unit and deposited on a cavity-specific basis. Only one closing unit (machine) is used here.

Two-stage process
In the two- or multi-stage process one of the components can even be made of metal with engineers currently endeavoring to replace metal parts as much as possible with technical plastics and hence to achieve a weight reduction. The two- or multi-stage process is performed by way of at least two closing units and hence on two autarkic machines or tools. Here too the insertion parts (metal) or premolded parts are inserted into the designated cavities of the respective mold and ejected by mold movement or deposited by a handling unit or a robot on a cavity-specific basis.

Process technologies
A distinction can basically be made in the process technologies between processes with separate contours, components or with contours that run into each other or components.

Composite injection molding
In the case of composite injection molding the combinations should exhibit good chemical adhesion to each other, such as for example a seal to technical components.

Assembly injection molding
Here material combinations that exhibit no chemical adhesion to each other are chosen, since otherwise the function of the component (e.g., a joint) is no longer assured.

Sandwich injection molding or co-injection technique
In this process different material properties are combined with each other, e.g., a rigid outer skin with a foamed core; therefore there does not necessarily have to be adhesion.

Multi-color injection molding
Here the same material is processed in different colors in one component (e.g., in multi-color car tail lights).

Fill simulation (mold flow analysis)
In all the processes mentioned and the design of molds, fill simulation is today one of the most important aids in the field of multi-component injection molding. It delivers detailed results and shows how problems can be avoided across the full injection-molding process. It provides useful information on pressure and temperature distribution and on possible positions of air pockets and joint lines in the component to be produced. In this way, potential flaws in molded parts and molds can in a fill simulation already be detected and eliminated in the development phase.

By detecting these potential flaw sources early, it is possible already at an early stage to eliminate the potential cost drivers in multi-component injection-molded parts.

Composite techniques
With the composite techniques there are a number of ways of combining the materials with each other. The correct composite technique depends here on a number of factors: Use of the component, choice of the materials used, mechanical load on the component, and quantities of the component required. According to these criteria the following composite techniques are currently predominant:
– mechanical bonding by undercuts,
– chemical bonding of the two materials,
– chemical bonding by previous priming (bonding agent),
– chemical bonding by surface treatment of the substrate.

Tool techniques
Rotary table
The rotary table is as a rule located between the closing-side mold half and the clamping plate. In this way it is possible to place components from a first cavity into a second cavity. First the material is injected into a first cavity (Figure 1). After intermediate opening and turning, the pre-molded part is now introduced into the second cavity. The cavity additionally opens a cavity for the second plastic. Following a successful injection process the mold is opened again and the finished two-component part is ejected. In parallel with this second operation the next pre-molded part is already produced in the first cavity.

Turning tool
With this turning technique the turning movement is effected not on the machine side, but by a mechanism in the mold (Figure 2). This can be performed for example by a plate in the mold.

Figure 1: Rotary-table process
a) Simultaneous filling of both cavities.
b) Finished part is ejected, turning of the rotary table on the machine plate including the pre-molded part.

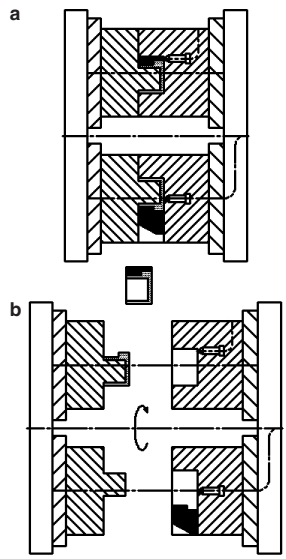
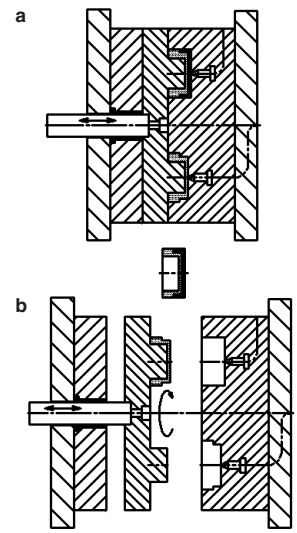


Figure 2: Turning tool
a) Simultaneous filling of both cavities.
b) Finished part is ejected, turning of the mold plate including the pre-molded part.



Turnstiles and turning cores
Instead of turning the complete closing-side mold half, in this solution only an intermediate plate in the mold (also called indexing plate) is turned with the pre-molded part.

Transfer technique or insertion technique
With the transfer or insertion technique the pre-molded part is transferred or inserted, via a gripper robot or by hand, from one cavity to the other (Figure 3). The second material is then injected here. Transfer technique also refers to the manufacture of the pre-molded part on a machine, removal and insertion in a second machine, and hence a two-stage process.

Core-back technique (slide technique)
The core-back technique (Figure 4) offers the opportunity of manufacturing two-component parts without having to open the mold. The first plastic is injected while the cavity for the second component is kept sealed by a sliding insert in the mold. Once this injection process has finished, the sealed cavity can be opened and the second component processed after the slide is actuated. On contact of the second material the contact-surface temperature can be monitored and adjusted much more effectively with this process than with rotating tool systems.

Figure 3: Transfer or insertion technique
a) Simultaneous filling of both cavities.
b) Finished part and pre-molded part are removed or inserted by the handling unit.

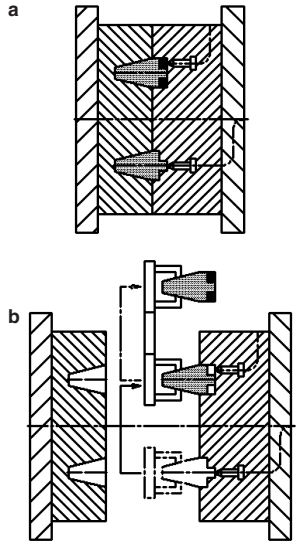
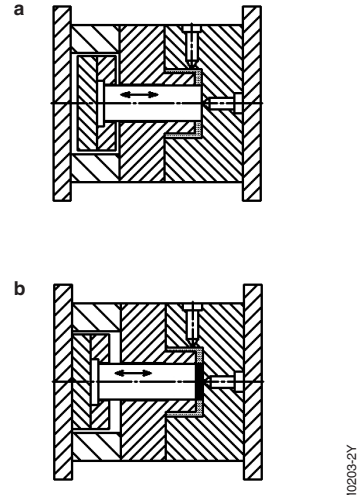


Figure 4: Core-back or slide technique
a) Cavity is filled with one component.
b) Core or slide is pulled and thus the sealed cavity is opened.



Rod and piston sealing systems

Hydraulic rod seals

Elements

A rod sealing system as a rule consists of five elements (Figure 1): Rod seal, rod guide, wiper, rod surface, and hydraulic fluid. The individual elements interact with each other. For example, a guide with a lot of play necessitates a seal which can elastically compensate this play. This must be observed particularly in a low-temperature application (–40 °C), since the flexibility of the elastomer decreases (see Elastomers). A further interaction can arise for example between the sealing element and the rod. If the rod surface is too rough, wear will occur on the sealing element, possibly on the wiper, and on the guide band. This wear results in leakage, even though the seal is not the cause of this.

Design

Rod seals are asymmetrical seals whose profile is optimized by a sealing edge in such a way that they can as effectively as possible wipe the medium to be sealed from the rod and hence provide a seal. The choice of sealing element is primarily dependent on pressure, required friction, temperature, media resistance, viscosity, lubricity, and price. The following general rule can be formulated: If the price and not the friction at medium pressures is the dominating factor and liquids of normal viscosity and lubricity are involved, elastomer grooved rings or polyurethane (PU) can be used. If, on the other hand, low friction or higher pressures are required, PTFE sealing-edge rings are to be recommended. The same applies in the case of high viscosity or poor lubricity.

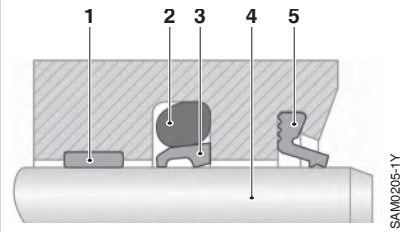
PTFE seals are activated by an elastomer energizing element – usually an O-ring – since PTFE has no rubber-elastic properties. It does however have very low friction and high dimensional stability. PTFE sealing-edge rings do not wipe the hydraulic fluid out of the rod-surface topography as well as elastomer and PU seals. They should for this reason exhibit sufficiently good back-pumping behavior, i.e. they should have the capacity to delivery a lubricant layer of a few molecule layers drawn out during the advance stroke back against the primary pressure into the pressure chamber during the return stroke. Thanks to this behavior, PTFE seals have the advantage of always running on a few molecule layers, which increases their service life. In contrast, PU and elastomers wipe the fluid from the surface in order to be tight. They exhibit only limited back-pumping capacity. On account of the lubricated extension and retraction and the low friction coefficient, the seal is prevented from squeaking or humming much more effectively than PU or elastomer seals.

Such systems are used for example in sealing the piston rod of a single-tube shock absorber or in sealing the rack of a hydraulic rack-and-pinion steering system. In this way, even the translatory movements of a positioning cylinder can be sealed in transmissions or in clutches.

Wipers

The function of wipers is to wipe dirt, dust and water from the rod and thus keep them away from the sealing system. Sealing-edge rings are used together with double wipers (with a wiper lip and an inward-wiping sealing lip) or with double-edge wipers (with a lip on which a wiper lip and a sealing edge are positioned). Double-edge wipers have a lower friction than double wipers. Grooved rings manage with a one-lip wiper. They constitute a compromise between a sharp wiper edge with good wiping properties and a

Figure 1: Hydraulic rod seal
1 Guide band, 2 O-ring, 3 Seal, 4 Shaft, 5 Wiper.



minimally rounded edge which does not wipe fluids as consistently.

Elastomers or PU can be used as materials for wipers. In exceptional cases PTFE can be used if substances permanently adhered to the rod are to be wiped. Even for wipers which do not come into contact with the hydraulic fluid, it is important to ensure media resistance on account of the draw-out lubricant layer of the rod seal (this exists for wipers – even PU and elastomers).

Rod guides

Guide rings – actually bands – must be designed in accordance with the requirements with regard to transverse force, friction, media resistance, and temperature. Temperature, friction and media resistance do not pose a problem to the most widely used extended PTFE guides. They cover a temperature range of –40 to +200 °C and have the lowest friction values. The media compatibility of PTFE is assured in most cases; only the resistance of the extenders has to be checked. These types of rings and bands can support medium transverse forces. For higher transverse forces a number of rings can be used or a thermoplastic (e.g., PA) can be chosen. It must be noted that all thermoplastic rings significantly lose their load capacity at temperatures in excess of 80 °C. The duroplastic bands familiar from industry are for price reasons only used in exception cases in the automotive sector (see Guides and guide bands).

Hydraulic fluids

Hydraulic fluids are for the most part specified. They should not contain any solid particles. These can reach between seal, guide, wiper or rod surfaces and damage or mark both the elements mentioned and the rod surface; this will then cause leakage. Lubricity and viscosity must be taken into account in the choice of seal.

Rod surface

A hardened or chromium-plated or even otherwise hard-coated surface (consultation with the seal manufacturer here) can at least reduce the problem of striation by hard particles. Lesser damage to PTFE seals normally repairs itself. The required roughness of the rod is dependent on the seal material and the relative speed. PTFE seals require a smoother surface than PU and elastomer seals. The value $R_z < 2.3$ µm applies to PU and elastomer seals with a maximum speed of less than 0.5 m/s. For PTFE seals this values is between 0.5 µm and 1.3 µm. The low value applies to high speeds and high frequencies (from approximately 0.8 m/s or 5 Hz), the higher value to slower movements. At high pressures and high speeds the surface should have a hardness of HRC > 50 with a case hardness depth $CHD > 0.15$. Generally, closed (plateau-shaped) surface profiles are to be preferred to open (pointed) profiles. All surface values are measured in the seal's axial direction of movement.

Installation of rod seals

Rod seals must be compressed in diameter in the course of installation if no split groove is provided. The compression limit is 13 % of the sealing diameter. Compression can be performed with a mounting tool (e.g., cone) with a lead-in chamfer of approximately 10 to 15° which upsets the sealing element to such an extent that it can pass through the bore diameter and snap into the groove.

A PU hose or an axially slotted tube can be used to push the seal into the cone. Subsequent calibration by a rod with a large diameter if possible and a lead-in chamfer of less than 15° are required at least for PTFE seals. The outside diameter of the lead-in chamfer should smaller than the seal diameter to be calibrated. The calibration is for the most part carried out with the original piston rod, on which a calibrating tip is attached. It is important here that it has a burr-free transition from the tip to the rod which has not sharp edges and thus does not damage the seal. The surface of the calibrating tip or the calibrating tool should have the same surface quality as that of the rod.

Installation by compression (kidney-shaped installation) of PTFE seals, as is sometimes suggested in industrial applications, must be avoided in automotive applications in that permanent deformation of the seals cannot be ruled out. A split groove must be provided if installation by compression is not possible.

Hydraulic piston seals

Seal, guide, medium, surface

The same applies by extension to piston sealing systems as to rod sealing systems. Normally, piston sealing systems have no wipers. This function is at least assumed in part by piston guide rings or piston bands. Piston seals are as a rule not subject to the same high tightness requirements ("only" internal leakage) as rod sealing systems, which must keep oil safely away from the atmosphere. Pistons seals are to exhibit equally good tightness in both directions of movement. For this reason, they have a symmetrical cross-section for the most part. Alternatively two grooved rings – back to back – must be used.

One interesting solution is a PTFE piston seal, which activates a sealing edge in each case as a function of the applied pressure. The pressure allows the seal to tilt in such a way that an edge is formed depending on the pressure direction and thereby can operate like a rod seal.

Piston guidance

The same principles as for rod sealing systems and rod guides apply to the choice of the right piston guide rings or guide bands. It proved to be advantageous to place a guide ring or guide band on either side (right and left) of the seal. This enables the guides to work to a limited extent as wipers (see Guides and guide bands).

Hydraulic fluids

See here Hydraulic fluids for rod sealing systems.

Piston surfaces

The same surface-roughness and surface-hardness figures apply to piston seals as to rod sealing systems. A non-hardened, drawn or rolled surface can also be used in special applications. Surface compression produces textures which allow the surface to become harder.

Installation of piston seals

Piston seals can be stretched via a cone with a lead-in chamfer of approximately 10 to 15° onto the piston diameter if the elongation does not exceed 30 % of the inside diameter. The elongated seals then snaps into the groove. A PU hose or an axially slotted installation tool (e.g., tube) can be used to push the seal into the cone.

Subsequent calibration is required for PTFE seals. This can be done with a separate calibrating tool or with the cylinder itself. The calibration tool is a cylinder with a lead-in chamfer of less than 15 ° and a bore diameter as close as possible to the piston diameter. The outside diameter of the lead-in chamfer should be at least as large as the outside diameter of the seal to be calibrated. Both in relation to the calibration tool and the lead-in chamfer on the cylinder it is important to ensure that the surface is as good as on the cylinder barrel so as to avoid damage to the sealing element. The transition from the calibration tool to the cylinder should be burr-free and have no sharp edges.

A split groove must be provided if the elongation of the seal exceeds 30 %.

Guides and guide bands

Guide bands have two essential functions to perform in rod and piston sealing systems: Guiding the piston or rod precisely and – if necessary – accommodating transverse forces. Different materials can be used to perform these functions, e.g., bronze, composite materials, or extended PTFE materials. Bronze bushings must be additionally lubricated and are expensive. In the case of composite materials, different guides and guide bands are possible. There are fabric-reinforced duroplastic guide rings, which support a relatively high transverse force, but are not suitable for dry running and even when lubricated have a relatively high friction. The market also offers guides which have a slotted sheet-metal ring on which bronze particles are sprayed and a PTFE layer is rolled in. These bushings offer outstanding guidance quality and low friction, but are at the expensive end of the spectrum.

A further guidance option is provided by wire-mesh bushings which are wrapped in PTFE and likewise have low friction.

Guide bands made of extended PTFE provide a popular and economical solution in vehicle manufacturing. They exhibit good friction, damp vibrations, and can support transverse forces up to a certain degree.

In the case of guides it must generally be noted that they do not assume the function of a seal. For this reason, the two ends of the guide rings (bands) create a gap (Figure 2). In the case of bushings or guides with narrow play (without a gap), it is important to ensure that no drag streams occur which are superimposed on the system pressure. A bypass may be required in these cases.

The thickness of guide bands depends on the diameter to be guided. The smaller the diameter, the harder it is to bend a thick band for example around a piston and hold it in position. Guide bands with thicknesses ranging from 0.85 to 2.5 mm are used in the automotive sector. The thinner the bands, the smaller their deformations under load and accordingly the better guidance quality they have. Refer to the relevant catalogs for calculation of the permissible load. PTFE materials extended to varying extents are available depending on application and medium.

Typical fields of application for guide bands are, for example, guiding the piston in a single-tube shock absorber or guiding the rod or the piston in a positioning cylinder for translatory movements (transmissions, clutches).

Figure 2: Guide band



Elastomer materials

General limits of use

The areas of application of elastomer materials are very wide-ranging (see also Materials, Elastomers). Correspondingly different also are the specific requirements which are imposed on the materials in each case (e.g., abrasion resistance, damping performance, insulating effect, long-lasting sealing force, resistance to fatigue cracks, compressive strength, media resistance).

Heat resistance and swelling behavior in oil

The following charts provide a general overview of the different limits of use in accordance with the general classification ASTM D2000 [15], based on the oil and heat resistance of the materials. Figure 1 shows the maximum working temperature and the change in volume after a storage period of 70 hours in the reference oil IRM 903. Figure 2 shows the operating-temperature range. These temperature ranges apply to applications in which contact with media which act aggressively against the respective material is excluded.

Elastomers

Generally speaking, the different elastomers can be roughly characterized as follows. The sequences of capital letters used for the elastomer ID codes are standardized in DIN ISO 1629 (Rubbers) [16].

IIR (isobutene-isoprene rubber)

Isobutene-isoprene rubber (butyl rubber) is characterized in particular by its very low permeability to air, water vapor and other gases. In addition, aside from good ozone, weather and aging resistance, IIR also has good resistance to organic and inorganic chemicals. The possible operating temperature is in the range –40 to +110 °C (briefly up to +120 °C).

Due to the low gas and moisture permeability IIR is used predominantly in tires.

CR (chloroprene rubber)

In general vulcanized chloroprene rubber exhibits relatively good ozone, weather, chemical and aging resistance, in addition high flame retardance, good mechanical properties, good abrasion resistance, and good flexibility at low temperature. The operating-temperature range is –35

Figure 1: Oil and heat resistance of elastomers

1 Value dependent on the acrylonitrile content (ACN)

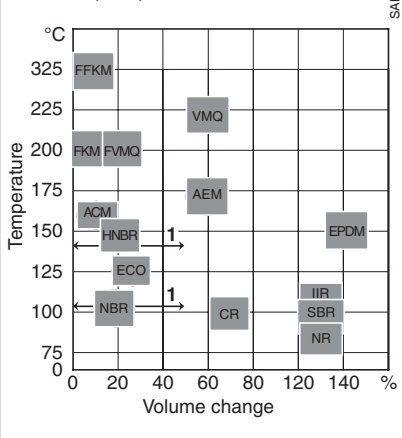
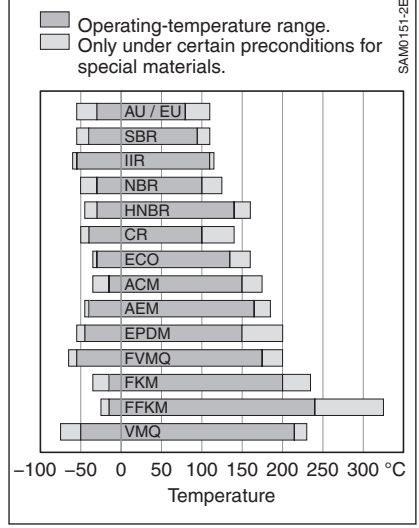


Figure 2: Operating-temperature ranges for elastomers



to +90 °C (briefly up to +120 °C). Special types can be used down to –55 °C.

CR materials are used among other things as seals against refrigerant, in outdoor areas, and in adhesive.

SBR (styrene butadiene rubber)
Of particular note in SBR materials are the outstanding mechanical properties, particularly the abrasion resistance. Accordingly, the material is mainly used in tires, but can also be found in V-belts.

In the field of seals SBR plays a lesser role, with resistance to typical media in an automobile such as oil or fuel not being given. The operating-temperature range is –40 to +110 °C (briefly up to +120 °C).

ECO (epichlorohydrin rubber)
The particularity of ECO materials is their good resistance to engine oils and fuels as used in automobiles. Moreover, the extremely low gas permeability, the weather and ozone resistance, and the low compression set must be mentioned.

The operating temperature is specified at –40 to +120 °C (briefly up to +140 °C), special types down to –60 °C.

Because of its outstanding dynamic properties, ECO is used in vehicles particularly for engine bearings and vibration dampers. But ECO is also frequently used in fuel lines.

NBR (nitrile butadiene rubber)
The properties of vulcanized NBR are mainly dependent on the ACN content (ACN: acrylonitrile), which can rang between 18 % and 50 %. This determines the oil resistance, which is however always contrary to the flexibility at low temperature.

Vulcanized NBR generally exhibits very good mechanical properties and good abrasion resistance at an operating temperature of –30 to +100 °C (briefly up to +120 °C). Special types can be used down to –50 °C.

NBR is used primarily for seals (also dynamic) or for hoses in contact with mineral oils and greases.

HNBR (hydrogenated nitrile butadiene rubber)
HNBR is produced by selective hydrogenation of the butadiene groups of NBR. In addition to the ACN content the properties are therefore also influenced by the degree of saturation set in the process.

The basic properties of HNBR are comparable with those of NBR. The operating temperature is however clearly improved by the saturation of the main chain and is in the range of –30 to +140 °C (briefly up to +160 °C). Special types can be used down to –40 °C.

They are used mainly in parts which come into contact with mineral oils and greases. However, the material also has good resistance to cooling water and to diesel fuel.

EPDM (ethylene propylene diene rubber)
EPDM has very good heat, ozone and aging resistance, also high elasticity and very good low-temperature performance. The operating temperature is with peroxidic cross-linking in the range –45 to +150 °C (briefly up to +175 °C). With sulfur cross-linking the heat resistance reduces to +130 °C (briefly up to +150 °C).

EPDM is frequently used in glycol-based brake fluids, in cooling water, in refrigerant, and, due to its outstanding weather resistance, for example in door seals.

ACM (polyacrylic rubber)
ACM exhibits very good ozone, weather and hot-air resistance, but only medium strength, low elasticity and relatively unfavorable low-temperature performance. Its operating-temperature range is –20 to +150 °C (briefly up to +175 °C). Special types can be used down to –35 °C.

ACM materials are used as seal materials primarily because of their particular resistance to lubricating oils with high additive levels (also sulfurous) in higher-temperature applications.

AEM (ethylene acrylate rubber)
Compared with ACM, AEM usually exhibits better mechanical properties, much better temperature resistance, and a very good compression set. Its operating-tem-

perature range is –40 to +160 °C (briefly up to +190 °C).

AEM materials are, like ACM, used as static seals in the underhood area (engine, transmission) primarily because of their particular resistance to lubricating oils with high additive levels in higher-temperature applications.

FKM (fluorine rubber)
Depending on their structure and fluorine content, fluorinated elastomers differ in their media resistance and flexibility at low temperature. Aside from their very good resistance to fuels, oils and other aggressive media, they are characterized by low gas permeability and very good ozone, weather and aging resistance. The operating-temperature range of fluorine rubbers is –20 to +200 °C (briefly up to +230 °C). Special types can be used down to below –40 °C.

FKM is used, because of its very good resistance to fuels, predominantly in applications such as fuel injectors, fuel lines, and the like.

It is important to mention the really high price of this elastomer group, which varies markedly depending on the polymer type used.

VMQ (silicone rubber)
The particularity of silicone rubbers is their high thermal resistance, the outstanding flexibility at low temperature, good dielectric properties, and above all good weather resistance. Silicone rubbers are hydrophobic and are used above all in the field of aqueous media. Special formulations are resistant to engine and transmission oils and water vapor. Depending on the version, the possible operating temperatures are in the range –50 to +200 °C (briefly in part also up to +230 °C).

Silicones are used in automobiles to seal electrical plug-in connections and sometimes also in radiators at high temperatures.

A subgroup, LSR (liquid silicone rubber), is very interesting because of its workability, above all for two-component parts with plastics.

FVMQ (fluorosilicone rubber)
Fluorosilicone rubber has, in comparison with silicones, much better chemical resistance in hydrocarbons, aromatic mineral oils and fuels. The possible operating-temperature range is specified at –55 to +175 °C (briefly in party also up to +200 °C).

PUR (polyurethanes)
The group of polyurethanes is extremely multilayered. The most diverse ranges of application can thus be individually covered, a standardization of the properties is not possible, and the materials are therefore conceived around their corresponding ranges of application. Depending on the recipe, outstanding recovery performance, extreme flexibility at low temperature, and optimum hydrolysis resistance are achieved. Generally, this material group exhibits low gas permeability, good mineral-oil resistance, high strength and abrasion resistance, but also more unfavorable heat resistance. Depending on the type, temperature operation ranges of –50 to +110 °C, briefly even higher, are achievable.

Characteristic data and testing of elastomer materials

Different properties are brought to bear depending on the use to which elastomers are put. To be able to assess the suitability of a material for a certain application, the custom is to conduct tests in advance which are meant to simulate the conditions in the later application. These tests are also intended to define or specify the properties of a material. Some of the most common test methods are described in the following.

Density

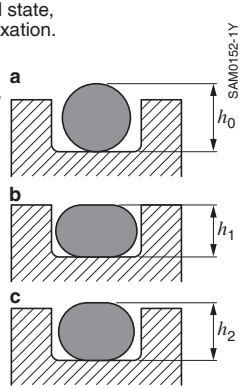
Density must be considered as one of the most fundamental properties. This test is – at a constant test temperature – to a large extent not dependent on further parameters such as parts geometry or similar, which is why it tends to be used as a specific material constant as the first identification check in material tests. Usually the density is determined using the buoyancy method according to ISO 1183-1 [17] in which the density is calculated from the measured weight in air and in a test liquid.

Compression set

An important parameter for sealing performance is the compression set of the elastomer material under the influence of temperature. Elastomers demonstrate under load, as well as an elastic component, a permanent, plastic deformation (Figure 3). The greater the compression-set value, the worse the elastic behavior of a material.

Figure 3: Depiction of compression set

a) Cylindrical disk before deformation,
b) in the deformed state,
c) after stress relaxation.
h_0 Original height,
h_1 Height in the deformed state,
h_2 Height after stress relaxation.



Common test specimens

The compression set is determined according to DIN ISO 815 [18]. A common test specimen is a cylindrical disk with a diameter of 13 mm and a height of 6.3 mm. The test can basically also be carried out with a suitable cross-section on seals (e.g., on O-rings).

Common parameters

The deformation is 25 %. The test temperature is chosen on the basis of the elastomer type. The test duration must be defined and is usually 24 h, 72 h, 168 h, or a multiple thereof.

To be able to compare the results, it is absolutely essential to define all the parameters exactly.

Calculation

The compression set is calculated as follows:

$$\text{Compression set} = \frac{h_0 - h_2}{h_0 - h_1}.$$

h_0 Original height,
h_1 Height in the deformed state,
h_2 Height after stress relaxation.

Compressive-stress relaxation

This test, for example according to 3384 [19], is similar to the compression-set test, but here it is not the recovery performance after stress relaxation that is measured but the specifically created counterforce in the compressed state that is determined. It requires much more technical apparatus to be measured, but has a much better correlation with the sealing force for static seals in the installed state.

Common test specimens

– Cylindrical disk with a diameter of 13 mm and a height of 6.3 mm
– O-rings with a diameter of 14 mm and a cross-section of 2.65 mm
– Plain compression rings with a diameter of 15 mm and a square cross-section of 2 mm edge length.

The test can basically be carried out with a suitable cross-section on any seal.

Common parameters
The deformation is 25 %. The test temperature is chosen on the basis of the elastomer type. The test duration must be defined and is usually 24 h, 72 h, 168 h, or a multiple thereof.

Test methods
Two different methods may be used: In method A the deformation and all the counterforce measurements are made at test temperature. In method B the test specimens are stored at the test temperature, with the deformation and all the force measurements being made at room temperature.

The measurement can be conducted in air and in media.

Calculation
The compressive-stress relaxation $R(t)$ is calculated as follows:

$$R(t) = \frac{F_0 - F_t}{F_0} \cdot 100$$

$R(t)$ Compressive-stress relaxation after a specified time t, expressed as % of the initial counterforce,
F_0 Initial counterforce after 30 min,
F_t Counterforce after the specified test duration t.

Hardness
The hardness is one of the most frequently mentioned properties of rubber materials. Nevertheless, the values can be very misleading. Hardness is the resistance of a body to the indentation of an indenting body of a specific shape under a defined compressive force.

Primarily two methods are used for hardness tests on standard test specimens and on finished parts made of elastomer materials:
– Shore hardness (indentation hardness according to the durometer method) according to ISO 7619-1 [20] for measurements on standard test specimens.
– IRHD hardness (International Rubber Hardness Degree) according to ISO 48 [21] for measurements on standard test specimens and finished parts.

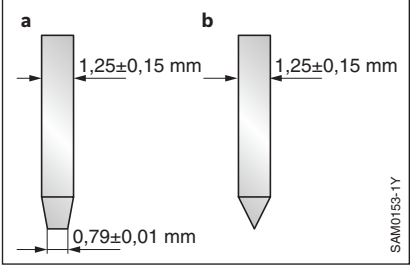
The measured values are dependent on
– the modulus of elasticity of the elastomer,
– the visco-elastic properties of the elastomer,
– the thickness of the test specimen,
– the geometry of the indenting body,
– the exerted pressure,
– the speed of the pressure increase,
– the time after which the hardness is registered.

The following also replies here: To be able to compare the results, it is absolutely essential to define all the parameters exactly. Because of these influencing variables, it is moreover not possible to compare the results obtained with a durometer (Shore hardness) directly with IRHD values.

The tests are to be carried out at 23 ± 2 °C.

Hardness tests according to Shore A and Shore D
The Shore A hardness tester (truncated cone) can be sensibly used in the hardness range 10 to 90. Harder specimens should be measured with the tester according to Shore D (pointed cone) (Figure 4).

Figure 4: Test specimens (indenting bodies) for hardness testing
a) Shore A, truncated cone,
b) Shore B, pointed cone.



Standard test specimen
– Diameter at least 30 mm
– Thickness at least 6 mm
– Top side and underside smooth and level.

For a thinner material it is permitted to use layers if the minimum specimen thickness is achieved with a maximum of three layers. None of the layers may be less than 2 mm thick.

Hardness tests according to IRHD
Testing the ball indentation hardness according to IRHD (International Rubber Hardness Degree) is applied both to standard test specimens and to finished parts.

The thickness of the test plate must be adapted to the hardness range. According to ISO 48 the hardness is subdivided into two ranges (Figure 5).

Soft: 10 to 35 IRHD
– Specimen thickness above 10 mm bis 12 mm.

Normal: Above 35 IRHD
– Specimen thickness 6 to 10 mm,
– Specimen thickness 1.5 to 2.5 mm (micro-hardness).

Hardness values determined on finished part or specimens of other dimensions deviate as a rule from the values m,measured on standard specimens. This applies mainly in the case of curved surfaces. In order to avoid measuring errors, it is essential to position the measuring tip on curved surfaces at the highest point. In the event of very small cord thicknesses, positioning aids such as magnifying glasses, centering pins and automatic laser positioning can be used.

Testing of tensile stress-strain properties according to DIN 53504 [22] or ISO 37 [23]
This test serves to determine the typical mechanical characteristic values of an elastomer such as tensile strength, elongation at break and the stress values of test specimens of a particular shape when subjected to stress at constant speed up to break (Figure 6).

Tensile strength
The tensile strength R_m is calculated as the quotient from the measured maximum force F_{max} and the initial cross-section A_0 of the test specimen:

$$R_m = \frac{F_{max}}{A_0}.$$

Elongation at break
The elongation at break ε_R is obtained as the quotient from the change measured at the moment of break $L_R - L_0$ (measured length L_R) and the initial measured length L_0 of the test specimen:

$$\varepsilon_R = \frac{L_R - L_0}{L_0}.$$

Figure 5: Test specimens (indenting bodies) for hardness testing acc. to IRHD
a) Soft, ISO 48 "L"/"CL",
b) Normal, ISO 48 "N"/"CNL",
c) Normal, ISO 48 "M"/"CM".

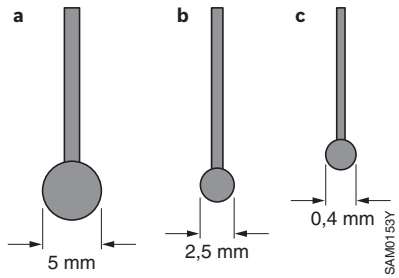
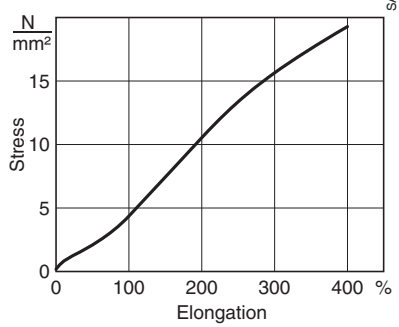


Figure 6: Force-deflection curve



Stress value
The stress value σ is the quotient from tensile force F_i available when a certain elongation is achieved and the initial cross-section A_0. Usually the value is specified at 100 %.

$$\sigma = \frac{F_i}{A_0}$$

Common test specimens
Common test specimens are S2 tensile bars which are stamped out of a 2 mm test plate (Figure 7).

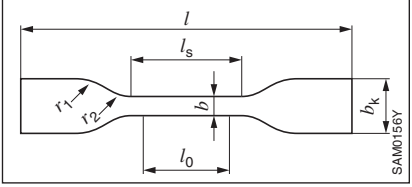
The test can basically also be carried out on O-rings and is described accordingly. However, the results determined here are not directly comparable with the values of standardized tensile bars.

Common parameters
The tests are generally carried out at the test temperature of 23 ± 2 °C. Often of interest are tests at increased or low temperature which are carried out in special tempering chambers.

Testing of tear-propagation strength according to DIN ISO 34-1 [24] or ASTM D624 [25]
This test measures the force that is required to tear a specified test specimen either by continuing an already existing cut or through the full width of the test specimen.

Common test specimens
Test specimens are tensile bars of different geometries which are stamped out of a 2 mm test plate. It must be noted that each test specimen leads to a completely different result, which is why only results of the absolutely identical method can be compared with each other.

Figure 7: Test specimen for testing the tensile stress-strain properties



Test temperature
The tests are carried out at the test temperature of 23 ± 2 °C.

Testing of media and temperature resistance
Elastomers, like all organic materials, have a limited service life. External factors such as media, temperature, ozone and oxygen influence the material properties. Elastomers can swell, shrink, harden or even crack if they are used under the wrong conditions.

In order to check the suitability of an elastomer under different conditions, it is necessary to conduct compatibility tests according to DIN ISO 1817 [26] in media or according to DIN 53508 [27] in air. Media compatibility can be tested with standardized test medium (e.g., ASTM oils, FAM test fuels) or in concrete operating media from the application.

The influence of a liquid on vulcanized elastomer can cause the following:
– absorption of the liquid by the elastomer,
– extraction of soluble constituents from the elastomer,
– chemical reaction with the elastomer.

The following changes are recorded in this test:
– change of mass, volume and dimensions,
– change of hardness and tensile stress-strain properties according to DIN 53504 [28] after exposure and if necessary after subsequent drying.

Common test specimens
S2 tensile bars according to DIN 53504 are used as test specimens. The test can basically also be carried out on seals (e.g., O-rings.

Common parameters
The test temperature is chosen on the basis of the elastomer type. The test duration must be defined and is usually 24 h, 72 h, 168 h, or a multiple thereof.

To be able to compare results, it is absolutely essential to define all the parameters exactly.

Testing of flexibility at low temperature
The properties of elastomers are very much influenced by temperature extremes. At high temperatures, the material may be subject to anything ranging from hardening to irreversible destruction. On the other hand, the material becomes rigid or hard and brittle at low temperature due to reduced chain mobility and loss of elasticity. This effect is reversible.

Different test methods are used to record this low-temperature effect:
– Calorimetric testing (DSC, Differential Scanning Calorimetry): This enables the T_g value (glass transition temperature) to be determined.
– Static measurement (TMA, Thermomechanical Analysis): This can be used to determine the TR10 value; it describes the recovery of an elongated specimen at low temperature by 10 %, according to ASTM D1329 [29] or ISO 2921 [30].
– Dynamic measurement (DMA, Dynamic-Mechanical Analysis): This can be used to determine the TR value (low-temperature characteristic for the sealing force at low temperature, see above) and further characteristic values.

Each test method leads to a slightly different measuring result, there is no direct correlation between the results.

As well as these analytical methods there is the possibility of determining the freezing-temperature range of a material by means of simple bending tests or a hardness test.

Testing of ozone resistance
Materials which in their polymer structure contain double bonds in the main chain (unsaturated polymers like NBR, HNBR and CR) are susceptible to ozone from the ambient air, above all in the elongated state. Testing of ozone resistance is conducted according to ISO 1431-1 [31] in special test chambers where the test specimens are subjected statically to tensile strain over the test duration. In this test it is assessed
– whether cracks are present in the test specimen after a defined test duration, or
– after what test duration cracks appear with a defined ozone concentration.

Common test specimens
The tests can be carried out on test bars or with a suitable cross-section also on seals.

To be able to compare results, it is essential to define clearly the following parameters: Temperature, ozone concentration, relative humidity, elongation, test duration, and test method.

Testing of rebound resilience
Elastomers are never 100 % elastic; instead, their property is always composed of a viscous component and an elastic component.

The damping property is of interest for certain applications. This can be determined for example by way of the rebound resilience according to DIN 53512 [32] or ISO 4662 [33], whereby a pendulum of defined mass (defined kinetic energy) strikes a test specimen and the height reached after the impact is measured.

Maximum damping corresponds to a value of 0 %. Minimum damping could be at a value of 100 %, but this cannot be achieved.

Common test specimens
The tests are carried out on 12 mm thick test specimens.

References

[1] ISO 3601: Fluid power systems – O-rings.
Part 1: Inside diameters, cross-sections, tolerances and designation codes.
Part 2: Housing dimensions for general applications.
Part 3: Quality acceptance criteria.
Part 4: Anti-extrusion rings (back-up rings).
Part 5: Specification of elastomeric materials for industrial applications.
[2] DIN 4768: Determination of values of surface roughness parameters R_a, R_z, R_{max} using electrical contact (stylus) instruments; concepts and measuring conditions.
[3] DIN EN ISO 4287: Geometrical Product Specifications (GPS) – Surface texture: Profile method – Terms, definitions and surface texture parameters.
[4] http://www.tss.trelleborg.com/de/de/service/design_support/oringcalculator_1/O-ringcalculator.html.
[5] http://www.tss.trelleborg.com/de/de/service/design_support/chemicalcompatibility_1/chemical-compatibility.html.
[6] DIN 443: Sealing push-in caps.
[7] DIN 3761: Rotary shaft lip type seals for vehicles.
Part 1: Terms; formula symbols, tolerances.
Part 2: Applications.
Part 3: Material requirements and methods of test.
Part 4: Visual irregularities.
Part 5: Test; measuring requirements and instruments.
[8] ISO 6194: Rotary shaft lip-type seals incorporating elastomeric sealing elements.
Part 1: Nominal dimensions and tolerances.
Part 2: Vocabulary.
Part 3: Storage, handling and installation.
Part 4: Performance test procedures.
Part 5: Identification of visual imperfections.
[9] DIN 3760: Rotary shaft lip type seals.
[10] DIN EN 10139: Cold rolled uncoated low carbon steel narrow strip for cold forming – Technical delivery conditions.
[11] DIN EN 10088: Stainless steels. Part 1: List of stainless steels.
Part 2: Technical delivery conditions for sheet/plate and strip of corrosion resisting steels for general purposes.
Part 3: Technical delivery conditions for semi-finished products, bars, rods, wire, sections and bright products of corrosion resisting steels for general purposes.
[12] VDMA 24317: Fluid power systems – Fire-resistant pressure fluids – Technical minimum requirements.
[13] DIN 24320: Fire-resistant fluids – Hydraulic fluids of categories HFAE and HFAS – Characteristics and requirements.
[14] DIN 51524: Pressure fluids – Hydraulic oils.
Part 1: HL hydraulic oils; minimum requirements.
Part 2: HLP hydraulic oils; minimum requirements.
Part 3: HVLP hydraulic oils; minimum requirements.

[15] ASTM D 2000: Standard Classification System for Rubber Products in Automotive Applications.
[16] DIN ISO 1629: Rubber and latices – Nomenclature.
[17] ISO 1183-1: Plastics – Methods for determining the density of non-cellular plastics.
Part 1: Immersion method, liquid pycnometer method and titration method.
[18] DIN ISO 815: Elastomers or thermoplastic elastomers – Determination of compression set.
Part 1: At ambient or elevated temperatures.
Part 2: At low temperatures.
[19] ISO 3384: Elastomers or thermoplastic elastomers – Determination of stress relaxation in compression.
Part 1: Testing at constant temperature.
Part 2: Testing with temperature cycling.
[20] ISO 7619-1: Elastomers or thermoplastic elastomers – Determination of indentation hardness.
Part 1: Durometer method (Shore hardness).
[21] ISO 48: Elastomers or thermoplastic elastomers – Determination of hardness (hardness between 10 IRHD and 100 IRHD).
[22] DIN 53504: Testing of rubber and elastomers – Determination of tensile strength at break, tensile strength at yield, elongation at break and stress values in a tensile test.
[23] ISO 37: Elastomers or thermoplastic elastomers – Determination of tensile stress-strain properties.
[24] DIN ISO 34-1: Elastomers or thermoplastic elastomers – Determination of tear strength.
Part 1: Trouser, angle and crescent test pieces.
[25] ASTM D 624: Standard Test Method for Tear Strength of Conventional Vulcanized Rubber and Thermoplastic Elastomers.
[26] DIN ISO 1817: Elastomers or thermoplastic elastomers – Determination of the effect of liquids.
[27] DIN 53508: Testing of rubber and elastomers – Accelerated ageing.
[28] DIN 53504: Testing of rubber and elastomers – Determination of tensile strength at break, tensile strength at yield, elongation at break and stress values in a tensile test.
[29] ASTM D 1329: Standard Test Method for Evaluating Rubber Property Retraction at Lower Temperatures.
[30] ISO 2921: Elastomers – Determination of low-temperature retraction (TR test).
[31] ISO 1431-1: Elastomers or thermoplastic elastomers – Resistance to ozone cracking.
Part 1: Static and dynamic strain testing.
Part 3: Reference and alternative methods for determining the ozone concentration in laboratory test chambers.
[32] DIN 53512: Testing of rubber and elastomers – Determination or rebound resilience (Schob pendulum).
[33] ISO 4662: Rubber – Determination of rebound resilience of vulcanized rubbers.

Detachable connections

Positive or form-closed joints

Operation

Positive or form-closed joints have the task of transferring forces, which themselves maintain contact via mating surfaces, by means of their geometrical shape. The forces are always transferred perpendicularly to the mating surfaces, resulting mainly in compressive and shear stress ([1], [2]).

Positive locking or form closure creates detachable connections, since as a rule a clearance or a transition fit exists (see Tolerances) between the touching surfaces (e.g. between a shaft and the bore of a hub, [3]). Depending on the choice of fit, relative axial movements may occur during operation. If necessary, they must be prevented by suitable locking devices. Among others, retaining rings according to DIN 471 [4] or slotted lock nuts according to DIN 981 [5] are used for this purpose.

Feather-key and Woodruff-key couplings

Feather-key couplings (Figure 1a) are used for the torsion-resistant connection of belt pulleys, gears, coupling hubs, etc. to shafts. Feather keys are sometimes used to secure frictional joints or to fix a specified position in the circumferential direction.

The cheaper Woodruff key (Figure 1b), the round side of which fits into the shaft, is mainly used for this purpose in automotive construction and to transfer smaller torques.

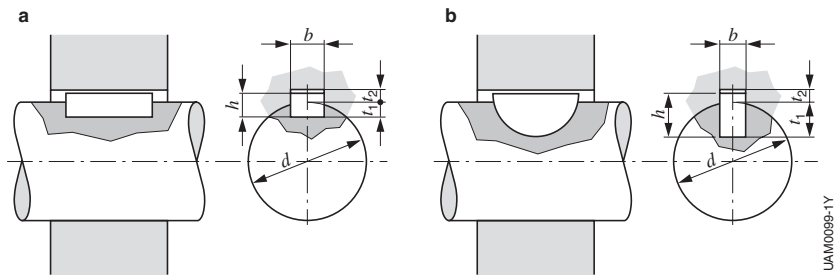
In the case of feather-key couplings, the groove faces lie against the feather-key faces. In contrast to keyed joints, there is clearance (backlash) between the rear of the feather key and the base of the groove. This means that the forces are transferred exclusively via the flanks of the feather key.

For feather-key width, the tolerance zone h 9 (key steel to DIN 6880 [6]) is provided. For groove widths b, the tolerance zones specified in Table 2 apply.

Table 1: Symbols and units

Quantity		Unit
D	Diameter	mm
F	Force	N
K_A	Application factor (service factor)	–
M_t	Torque	Nm
b	Width	mm
d	Diameter	mm
h	Height	mm
i	Number of shear surfaces	–
l	Length	mm
l_{tr}	Supporting feather-key length	mm
n	Number of drivers	–
p	Surface pressure	N/mm²
t_1	Groove depth (shaft)	mm
t_2	Groove depth (hub)	mm
σ_b	Bending stress	N/mm²
τ_s	Shear stress	N/mm²
φ	Contact-surface ratio	–

Figure 1: Positive or form-closed joints
a) Feather-key coupling, b) Woodruff-key coupling. Symbols, see Table 1.



A sliding fit must be used if a hub must move on the shaft in the longitudinal direction (e.g. gear in manual-shift transmission). Usually, the sliding spring is firmly bolted into the shaft groove. Round (Shape A) and angular (Shape B) feather keys (Figure 2) are manufactured. DIN 6885 [7] defines the standard in terms of their shape and dimensions, depending on the shaft diameter (Table 3).

In practice, feather keys are designed only for surface pressure. If $p \leq p_{perm}$, the required supporting feather-key length (see Table 1) is

$$l_{tr} = \frac{2\,K_A\,M_t}{d\,(h - t_1)\,n\,\varphi\,p_{perm}} \quad .$$

For round-faced feather keys (Shape A), the feather-key length is $l = l_{tr} + b$. For straight-faced ones (Shape B), it is $l = l_{tr}$. For the permissible surface pressures, the standard gives $p_{perm} = 0.9\,R_{e,min}$, where $R_{e,min}$ is the minimum yield point of the shaft, hub or feather-key material. With one feather key ($n = 1$), the contact-surface ratio is $\varphi = 1$, and with two feather keys, it is $\varphi = 0.75$.

Table 2: Tolerances for groove widths

Groove fit	Fixed fit	Easy fit	Sliding fit
in hub	*P* 9	*N* 9	*H* 8
in shaft	*P* 9	*J* 9	*D* 10

Profiled shaft-hub connections

Instead of inserting multiple feather keys into shaft grooves, the shaft cross-section can also be shaped directly in the form of a polygonal profile, and the mating hub cross-section has a corresponding shape. (Table 4). Profiled shafts are also used if the hub must be axially movable relative to the shaft (e.g. steering column for a height-adjustable steering wheel). The profiled shaft connection has the advantage

Figure 2: Feather-key shapes
a) Shape A, b) Shape B.
Symbols, see Table 1.

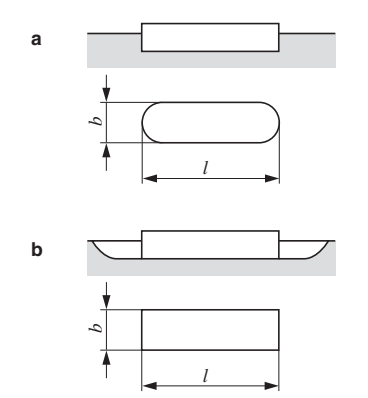


Table 3: Feather-key dimensions according to DIN 6885

Shaft diameter *d* over mm	Shaft diameter *d* up to mm	Width × height *b* × *h* mm	Groove depths *t*1 mm	Groove depths *t*2 mm	Length *l* mm
6	8	2 × 2	1.2	1.0	6…20
8	10	3 × 3	1.8	1.4	6…36
10	12	4 × 4	2.5	1.8	8…45
12	17	5 × 5	3.0	2.3	10…56
17	22	6 × 6	3.5	2.8	14…70
22	30	8 × 7	4.0	3.3	18…90
30	38	10 × 8	5.0	3.3	22…110
38	44	12 × 8	5.0	3.3	28…140
44	50	14 × 9	5.5	3.8	36…160
50	58	16 × 10	6.0	4.3	45…180
58	65	18 × 11	7.0	4.4	50…200
65	75	20 × 12	7.5	4.9	56…220
75	85	22 × 14	9.0	5.4	63…250
85	95	25 × 14	9.0	5.4	70…280
95	110	28 × 16	10.0	6.4	80…320
Feather-key lengths in mm:		6, 8, 10, 12, 14, 16, 18, 20, 22, 25, 28, 32, 36, 40, 45, 50, 56, 63, 70, 80, 90, 100, 110, 125, 140, 160, 180, 200, 220, 250, 280, 320			

that it does not require any additional separator (feather key) to transfer torque. The hub can be centered on the shaft over the smallest shaft diameter. This is called internal centering and very smooth running can thus be achieved.

The hub can also be centered via the flanks of the drivers or the toothing. This flank centering ensures very low circumferential backlash between shaft and hub and is therefore very suitable for alternating and jerky torques. As with the feather key, a rough design is drawn up based on surface pressure.

Bolt and pin connections

Bolt and pin connections provide a simple and inexpensive means of connecting two or more components. These are among the oldest and most widely used types of connection.

Bolt connections

Bolt connections are mainly used for joining linkages (Table 5), shackles, chain links, and connecting rods, as well as axles for bearing impeller rings, rollers, levers, etc. Since relative movements occur in these connections, at least one part must be movable. The main prevailing stresses are surface pressure (Table 6) and shear. Bending stress is negligible in most cases. It only occurs to a significant extent in the case of bolt connections which are relatively long in relation to their diameter.

Pin connections

Pins are suitable for the permanent connection of hubs, levers, and set collars on shafts or axles (e.g. transverse-pin joint). They also secure the precise position of two machine parts, and are suitable as guide pins to fix springs, etc. (Table 5). Since they are forced into holes as press fits with tolerances, all parts are permanent. The pin can be knocked out with a punch without damaging the components.

Table 4: Profiled shaft-hub connections

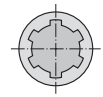
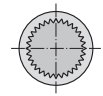
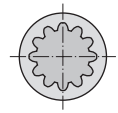
Designation	Standard	Illustration	Driver	Centering	Contact-surface ratio
Spline shaft	ISO 14 [8] DIN 5464 [9]	 UAM0101-1Y	Prismatic driver	Inner	$\varphi = 0.75$
				Flanks	$\varphi = 0.9$
Toothed shaft with grooved toothing	DIN 5481 [10]	 UAM0101-2Y	Grooved toothing	Flanks	$\varphi = 0.5$
Toothed shaft with involute toothing	DIN 5480 [11] DIN 5482-1 [12]	 UAM0101-3Y	Involute toothing	Flanks	$\varphi = 0.75$

Table 5: Bolt and pin connections

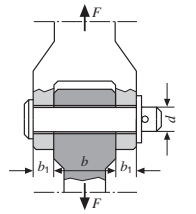
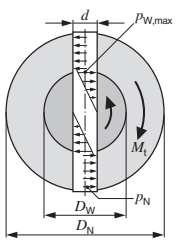
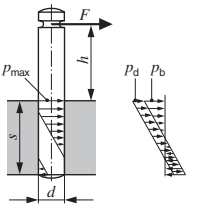
Designation	Illustration	Calculations
Articulated joint		Surface pressure in fork: $p_G = \frac{F}{2\, b_1\, d} \le p_{perm}$ Surface pressure in rod: $p_S = \frac{F}{b\, d} \le p_{perm}$ Surface pressure in pin: $\tau_S = \frac{4\, F}{i\, \pi\, d^2} \le \tau_{S,perm}$
Transverse-pin joint		Surface pressure in shaft: $p_{W,max} = \frac{6\, M_t}{d\, D_W^2} \le p_{perm}$ Surface pressure in hub: $p_N = \frac{4\, M_t}{d\,(D_N^2 - D_W^2)} \le p_{perm}$ Surface pressure in pin: $\tau_S = \frac{4\, M_t}{D_W\, \pi\, d^2} \le \tau_{S,perm}$
Guide pin		Maximum pressure: $p_{max} = p_b + p_d = \frac{F}{d\, s}\left(1 + 6 \cdot \frac{h + s/2}{s}\right) \le p_{perm}$ Bending stress at clamping point: $\sigma_b = \frac{32\, F\, h}{\pi\, d^3} \le \sigma_{b,perm}$ Shear stress in clamping point: $\tau_s = \frac{4\, F}{\pi\, d^2} \le \tau_{S,perm}$

Table 6: Permissible mean surface pressure for bolt and pin connections

Permanent fits			Sliding fits	
Material	Mean surface pressure		Material pairing	Mean surface pressure
	static	swelling		
	p_{perm} N/mm²	p_{perm} N/mm²		p_{perm} N/mm²
Gray cast iron	70	50	St / gray cast iron	5
S 235 (St 37)	85	65	St / CS	7
S 295 (St 50)	120	90	St / Bz	8
S 335 (St 60)	150	105	St hard. / Bz	10
S 369 (St 70)	180	120	St hard. / St hard.	15

Frictional joints

Operation

For frictional joints, press fits are produced in the joints (the friction surfaces being the effective areas) in which the parts to be assembled are in direct contact (Figure 1). The surface pressure p can be generated by bolt forces, keys, elastic separators, or the elasticity of the components themselves. The resulting normal force $F_N = pA$ (with friction surface A) induces a frictional force F_R, which opposes a movement caused by external forces ([1], [2]).

Press fit

Application

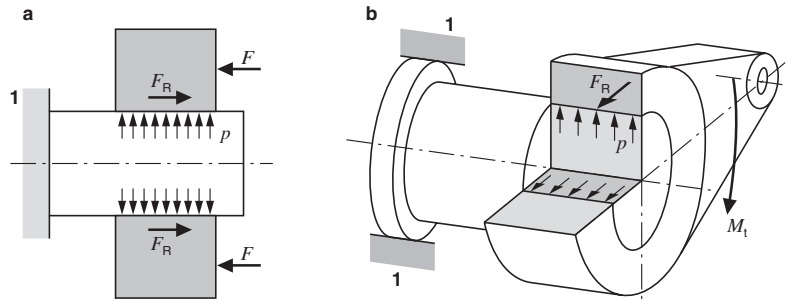
For a press fit (cylindrical interference fit), the required surface pressure is produced by the elastic deformation of the shaft and hub, resulting from an interference fit. "Interference fit" is the pairing of cylindrical fitting parts which have an interference before assembly (Figure 2).

Because press fits are easy to produce and can transfer even jerky and variable torques and linear forces, they are suitable for the joints of cylindrical surfaces which do not have to be released afterwards (e.g. gear on shaft, wheel on axle, bushing in housing). Cylindrical press fits are classed as hard-to-detach connections, since with high interference during removal the surfaces of shaft and bores are detrimentally affected.

Table 1: Symbols and units.

Quantity		Unit
A	Area (friction surface)	mm^2
C	Taper ratio	–
D	Diameter	mm
E	Modulus of elasticity	N/mm^2
F	Force	N
F_a	Axial force	N
F_N	Normal force	N
F_R	Frictional force	N
K_A	Application factor (service factor)	–
M_t	Torque	Nm
$M_{t,\,nomn}$	Nominal load torque	Nm
Q	Diameter ratio	–
R_e	Yield point	N/mm^2
R_m	Ultimate strength	N/mm^2
R_z	Surface roughness	mm
S_B	Factor of safety against rupture	–
S_F	Factor of safety against yield	–
U	Tolerance	mm
Z	Allowance	mm
b	Hub width	mm
d	Diameter	mm
l	Taper or lever length	mm
n	Number of bolts	–
p	Surface pressure	N/mm^2
t	Temperature in Celsius	°C
α	Taper angle	°
α_A	Coefficient of linear expansion, outer part	$10^{-6}/K$
α_i	Coefficient of linear expansion, inner part	$10^{-6}/K$
μ	Coefficient of friction	–
ν	Poisson's ratio	–
ξ	Specific allowance	mm^3/N
σ_{perm}	Permissible stress	N/mm^2

Figure 1: Frictional joints.
a) Axially stressed joint,
b) Tangentially stressed joint.
1 Gripped side of joint.
F Force, F_R Frictional force, p Surface pressure, M_t Torque.



Elastic design of cylindrical interference fits
A press fit must be designed so that at least a minimum surface pressure p_{min} is present in order to ensure the transfer of the greatest occurring stress, and so that a maximum surface pressure p_{max} is not exceeded and the components are not overstressed.

Basically, two calculation objectives are possible: The required fit is defined for a given load (Table 2) or the permissible load is determined for a given fit (Table 3). In the following equations index I denotes the inner part (shaft) and index A the outer part (hub) of the press fit.

The diameter ratios (see Figure 2)

$$Q_A = \frac{D_F}{D_{Aa}} \text{ and}$$

$$Q_I = \frac{D_{Ii}}{D_F}$$

and the specific allowance [1]

$$\xi = D_F \left[\frac{1}{E_I} \left(\frac{1+Q_I^2}{1-Q_I^2} - \nu_I \right) + \frac{1}{E_A} \left(\frac{1+Q_A^2}{1-Q_A^2} + \nu_A \right) \right]$$

can be used to design press fits with reference to their function and the required component safety.

The nominal diameter is inserted into the calculation for the joining diameter D_F. When joining an interference, loss results from plastic leveling of the roughness

Table 2: Definition of fits for given stress [1]

Stress M_t or F_a given	$\sigma_{perm} = R_e / S_F$ or $\sigma_{perm} = R_m / S_B$ given
Required fit: $p_{min} = \frac{\sqrt{F_a^2 + \frac{4 M_{t,nom}^2}{D_F^2}}}{\mu \pi D_F b} K_A$	Permissible fit in hub: $p_{max} = \left(1 - Q_A^2\right) \sigma_{perm} / \sqrt{3}$ Permissible fit in hollow shaft: $p_{max} = \left(1 - Q_I^2\right) \sigma_{perm} / \sqrt{3}$
Required allowance: $Z_{min} = p_{min}\, \xi$	Permissible allowance: $Z_{max} = p_{max}\, \xi$
Required interference: $U_{min} = Z_{min} + 0.8\,(R_{zI} + R_{zA})$	Permissible interference: $U_{max} = Z_{max} + 0.8\,(R_{zI} + R_{zA})$
Select ISO fits with $U_k \geq U_{min}$ and $U_g \leq U_{max}$	

Figure 2: Press fits
a) After joining,
b) Before joining.
1 Inner part (shaft), 2 Outer part (hub).
F_R Frictional force, p Surface pressure, M_t Torque, b Hub width, D_F Joining diameter,
D_{Ia} Shaft outside diameter, D_{Ii} Shaft inside diameter,
D_{Aa} Hub outside diameter, D_{Ai} Hub inside diameter.

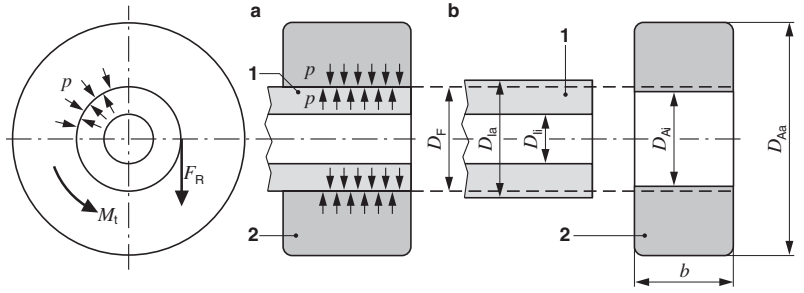


Table 3: Determination of stress for given fit

Smallest interference U_k given	Greatest interference U_g given
Smallest allowance: $Z_k = U_k - 0.8\,(R_{zI} + R_{zA})$	Greatest allowance: $Z_g = U_g - 0.8\,(R_{zI} + R_{zA})$
Smallest pressure: $P_k = \frac{Z_k}{\xi}$	Greatest pressure: $P_g = \frac{Z_g}{\xi}$
Permissible stress: $M_t = 0.5 p_k \mu \pi D_F^2 b$ (only M_t) $F_a = p_k \mu \pi D_F b$ (only F_a) $M_t = \frac{D_F}{2}\sqrt{(p_k \mu \pi D_F b)^2 - F_a^2}$ (F_a given) $F_a = \sqrt{p_k \mu \pi D_F b - \frac{4M_t}{D_F^2}}$ (M_t given)	Hub safety factor: $S_F = \frac{1-Q_A^2}{\sqrt{3}\,p_g} R_e$ or $S_B = \frac{1-Q_A^2}{\sqrt{3}\,p_g} R_m$ Hollow-shaft safety factor: $S_F = \frac{1-Q_I^2}{\sqrt{3}\,p_g} R_e$ or $S_B = \frac{1-Q_I^2}{\sqrt{3}\,p_g} R_m$

peaks. According to DIN 7190 [13], the smoothing of the two surfaces with 40 % of the averaged peak-to-valley heights R_{zI} (shaft) and R_{zA} (bore) is taken into account in the calculation.

The maximum tensions occur according to the equations in Table 2 and Table 3 at the internal diameters of hollow shaft and hub. Solid shafts are noncritical, and do not usually have to be calculated.

Assembly

There are two types of press fit depending on the assembly method: linear and transverse. Linear press fits are produced by "cold" assembly at room temperature. The large press-fit forces required are usually applied by hydrostatic presses. The press-fit velocity should not exceed 2 mm/s. For the press-fit force:

$$F_e = \frac{(U_g - 0.8\,(R_{zI} + R_{za}))\,\mu \pi D_F b}{\xi}.$$

Before transverse press fits are made, either the outer part is expanded by heating, or the diameter of the inner part is reduced by supercooling, so that the parts can be assembled stress-free. If the outer part is heated, it shrinks onto the

Table 4: Poisson ratio, modulus of elasticity, and linear coefficient of thermal expansion for metallic materials

Material	Poisson ratio ν	Modulus of elasticity E [N/mm^2]	Coefficient of linear expansion α [10^{-6}/K]	
			Heating	Cooling
Gray cast iron	0.24	100,000	10	–8
Malleable cast iron	0.25	90,000 to 100,000		
Steel	0.3	200,000 to 235,000	11	–8.5
Bronze	0.35	110,000 to 125,000	16	–14
Red brass	0.35 to 0.36	110,000 to 125,000	17	–15
CuZn	0.36	80,000 to 125,000	18	–16
$MgAl_8Zn$ AlMgSi	0.3 0.34	65,000 to 75,000	23	–18

inner part as it cools (shrink fit). If the inner part is cooled so that it expands when it is heated to room temperature, this is an expansion fit. To make force-free assembly possible, a joint clearance of $ED = 0.001 \cdot D_F$ must be provided.

With the ambient temperature t_U, the maximum interference U_g and the coefficient of linear thermal expansion α (Table 4) the assembly temperature of the hub comes out for a shrink fit at

$$t_A = t_u + \frac{U_g + \Delta D}{\alpha_A D_F}$$

and for an expansion fit at

$$t_I = t_u - \frac{U_g + \Delta D}{|\alpha_I|\, D_F}.$$

Tapered connection

Application

A tapered connection (tapered interference fit) is suitable for transferring dynamic forces and torques. They are used predominantly to secure hubs to shaft ends, e.g. the V-belt pulley on the alternator or the drill head in the drill spindle. Taper and taper angle are standardized, among others also connecting elements like the terminals on starter batteries, in DIN 254 [14].

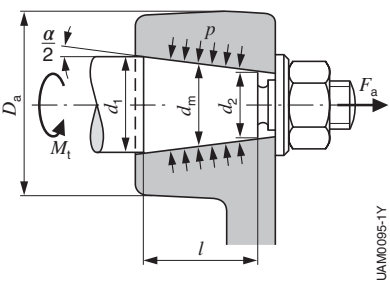
Tapered connections for the most part can be retightened and easily detached. As a shaft-hub connection they exhibit no shaft weakening and have very good centering (i.e. no imbalance).

However, the disadvantages are the high production costs and the lack of adjustability in the axial direction. The following applies (Figure 3) to the

taper ratio $C = \frac{(d_1 - d_2)}{l}$
and to the
taper angle $\tan\frac{\alpha}{2} = \frac{(d_1 - d_2)}{2l}$.

The following taper ratios are specified as guidelines (DIN 254 [14], DIN 406 [15]):
$C = 1:5$ connection is easily detachable,
$C = 1:10$ connection is detachable only with difficulty.

Figure 3: Tapered connection
D_a Hub outside diameter,
d_1 Large taper diameter,
d_2 Small taper diameter,
d_m Middle taper diameter,
l Common length,
p Surface pressure,
α Taper angle,
M_t Torque,
F_a Assembling force (fastener force).



Operation

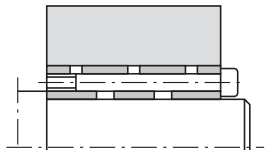
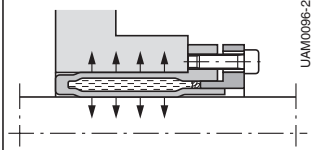
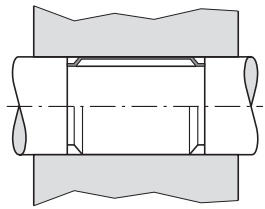
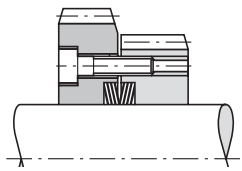
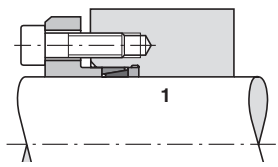
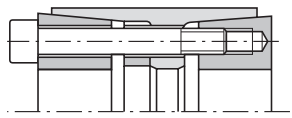
The effective area pair of a tapered connection takes the form of a truncated cone (frustum). The required surface pressure p is usually applied by an axial bolt force F_a. The relationship between the axial press-fit force F_a and the transferable torque M_t is expressed by the following equation [1]:

$$F_a \geq \frac{2K_A M_{t,nom}}{\mu_u d_m}\left(\sin\frac{\alpha}{2} + \mu_a \cos\frac{\alpha}{2}\right).$$

Load excesses (impacts) which can occur during operation are taken into consideration with the application factor K_A, which according to DIN 3990 [16] can range between 1 and 2.25. This equation also takes into account the possible difference between the coefficients of friction μ_u in the circumferential direction and μ_a in the axial direction. Self-locking exists if an expulsion force (F_a becomes negative) is required to detach the connection. This means that a torque can be transferred even if the bolt is removed after the parts are axially stressed. In contrast, with a non-self-locking tapered connection, there is no pressure between the effective areas after the axial application force is released. The condition for self-locking is:

$$\frac{\alpha}{2} \leq \arctan\mu_a.$$

Table 5: Taper-lock joints.

Designation	Illustration	Features
Clamping sleeve (Spieth)	 UAM0096-1Y	Axial distortion is applied to enlarge the outside diameter of the clamping sleeve and reduce the inside diameter. The danger of loosening under dynamic loads is reduced by using long bolts.
Hydraulic hollow-jacket spring collet (Lenze)	 UAM0096-2Y	Axial distortion is applied to generate a pressure in a thin-walled hollow cylinder. The spring collet is self-centering because of even pressure distribution, and thus has true-running properties. At higher temperatures, the thermal expansion of the pressure fluid must be taken into account.
Tolerance ring (Dr Tretter)	 UAM0096-3Y	Tolerance rings are slotted and made from thin wave-shaped sheet metal. The required initial force is provided by forced deformation of the elastic connecting member. They can bridge relatively large machining tolerances, compensate for thermal expansion, and transfer torques.
Star washer (Ringspann)	 UAM0096-4Y	Star washers are thin-walled, very flat conical shells with radial slots. The axial initial force is translated by forced deformation into a fivefold to tenfold radial force. Star washers do not center.
Taper-lock ring (Ringfeder)	1 Precentering  UAM0096-5Y	Taper-lock rings consist of two coaxially arranged conical rings. Radial prestressing is achieved by axially distorting the rings. Taper-lock rings do not center. They are not self-locking and thus can easily be detached.
Taper-lock set (Bikon)	 UAM0096-6Y	Axial distortion is applied by means of the bolts belonging to the taper-lock set. Taper lock sets are very true-running and can transfer very large torques, particularly with multiple pairs of effective areas.

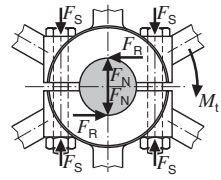
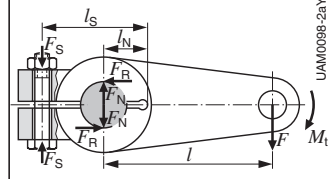
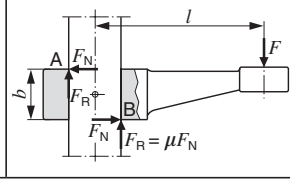
Component safety
The critical component is the hub. It is calculated as an open, thick-walled hollow cylinder. If $Q = d_m/D_a$, the following applies to the safety of the hub, with elastic design, depending on the modified shear-stress theory (DIN 7190):

$$S_F = \frac{1-Q^2}{\sqrt{3}} \cdot \frac{\left(\sin\frac{\alpha}{2}+\mu_A\cos\frac{\alpha}{2}\right)\pi d_m l}{F_{a,max}}.$$

Taper-lock joints
The required surface pressure in the effective areas can also be applied by elastic separators (Table 5). The great advantage of these taper-lock joints is that they can be used to fix hubs, gears, couplings, etc. securely to smooth, cylindrical shafts. Unlike cylindrical interference fits, they can be freely adjusted axially and tangentially, and are above all easily detachable. They are, therefore, particularly suitable for hubs (e.g. belt pulleys), which must be adjustable and replaceable. Their disadvantages include the space required and the high costs. They are usually designed as specified by the manufacturer (see manufacturer catalogs and Table 5).

Clamp joints
In the case of clamp joints, external forces apply the required surface pressure to the joint, mostly by means of bolts. The joints with a sectional or slotted hub are preferably used for low torques which have little fluctuation. Their advantage is that the hub position is easily adjustable in the axial and tangential directions. They provide a very easy means of fixing wheels or levers to smooth shafts. However, there are also self-locking clamp joints. Here, the tilting force F_K generates edge pressures in A and B (Table 6) to prevent axial movement.

Table 6: Clamp joints

Designation	Illustration	Features
Clamp joint with separated hub	 UAM0098-1aY	Transferable torque: $M_t = nF_S\mu\frac{\pi}{2}D_F$. n Number of bolts, F_S Fastener force, F_R Frictional force, F_N Normal force, M_t Torque.
Clamp joint with slotted hub	 UAM0098-2aY	Transferable torque: $M_t = nF_S\mu\frac{\pi}{2}D_F\frac{l_s}{l_N}$. n Number of bolts, F Force, F_S Fastener force, F_R Frictional force, F_N Normal force, M_t Torque, l_S Bolt lever arm, l_N Hub-center lever arm.
Self-locking clamp joint (screw-clamp principle)	 UAM0098-3aY	Condition for self-locking: $\frac{l}{b} \geq \frac{1}{2\mu}$. F Force, F_R Frictional force, F_N Normal force, μ Coefficient of friction, b Hub width, A, B Contact points.

Keyed joints

Longitudinal keyed joints

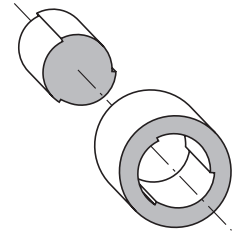
A one-sided radial distortion is achieved by driving in a standardized key (key angle = 0.57°) between the shaft and hub. However, because of their imprecise assembly (hammer assembly) and the resulting eccentricities, keys are only of secondary importance.

Circular keyed joint

A new type of keyed joint is the 3-part circular-key sectional profile (Figure 4). Three circular keys are arranged circumferentially on the cylindrical surface of a shaft (inner part). The hub (outer part) contains an appropriate number of corresponding keys in a cylindrical hole. Twisting results in radial distortion and this can transfer large axial and tangential forces in any direction.

As opposed to press fits, circular keyed joints are detachable. They are used, for example, for shaft-hub connections, for camshafts and for hinges in vehicle construction.

Figure 4: 3-part circular-key sectional profile



Threaded fasteners

Basic principles

Operation

A threaded fastener is used to make secure joints that are detachable any number of times. The purpose of screws and bolts is to stress the mating parts so that the static or dynamic operating forces acting on the joint do not cause any relative movement between the parts ([1], [17], [18], [19], [20]).

When a fastener is tightened or loosened, a screwing motion takes place. In one full rotation of the fastener, an axial shift corresponding to the pitch P occurs. If a fastener line is uncoiled on a cylinder that has a flank diameter d_2, it will produce a straight line with a pitch angle $\varphi = \arctan(P/(\pi d_2))$.

In general, fasteners are right-threaded (the straight line rises to the right). Special applications may require left-threaded bolts.

For normal fixing bolts or screws, metric thread profiles (DIN 13 [21], ISO 965 [22]) are used. For pipes, fittings, threaded flanges, etc., pipe threads (DIN ISO 228-1 [23], DIN EN 10226-1 [24]) are used. Refer to Figures 7 through 9 and Tables 7 through 10 for the dimensions of these threads.

Table 1: Symbols and units

Quantity		Unit
A	Cross-sectional area	mm²
A_S	Stress area	mm²
D_{Km}	Effective diameter for friction torque in fastener-head or nut bearing surface	mm
E	Modulus of elasticity	N/mm²
F_A	Axial operating force	N
F_K	Clamping force	N
F_M	Assembly preload	N
F_N	Normal force	N
F'_N	Normal-force component in plane force polygon	N
F_{PA}	Additional plate force	N
F_Q	Transverse force, operating force applied perpendicular to fastener axis	N
F_U	Peripheral force	N
F_S	Fastener force	N
F_{SA}	Additional fastener force	N
F_V	Preload	N
F_z	Loss of preload due to settling	N
M_A	Tightening torque	Nm
M_G	Effective tightening-torque component in thread	Nm
M_{KR}	Head friction torque	Nm
M_L	Release torque	Nm
P	Thread pitch	mm
R_e	Yield point	N/mm²
$R_{p0.2}$	0.2 % yield strength	N/mm²
R_P	Spring rate of stressed parts (plates)	N/mm
R_S	Spring rate of fastener	N/mm
R_z	Surface roughness	µm
W_t	Section modulus against torsion	mm³
d	Nominal thread diameter	mm
d_2	Thread flank diameter	mm
d_3	Thread root diameter	mm
d_h	Hole diameter of stressed parts	mm

Quantity		Unit
d_w	Outside diameter of flat fastener-head or nut bearing surface	mm
f_A	Elastic linear expansion by F_A	mm
f_{PV}	Elastic linear expansion of stressed parts by F_V	mm
f_{SV}	Elastic linear expansion of fastener by F_V	mm
f_z	Contact stress	mm
i	Friction-surface pairs	–
l	Length	mm
m	Nut height or thread reach	mm
n	Force-application factor	–
n_S	Number of fasteners	–
α	Flank angle of thread	°
α_a	Tightening factor	–
μ_G	Coefficient of friction in thread	–
μ_K	Coefficient of friction in head bearing surface	–
μ_T	Coefficient of friction in parting line	–
μ'_G	Apparent coefficient of friction in thread	–
ρ'_G	Angle of friction to μ'_G	–
σ_a	Alternating cyclic stress on fastener	N/mm²
σ_a	Permissible variable stress	N/mm²
$\sigma_{red,B}$	Equivalent stress in operating state	N/mm²
$\sigma_{red,M}$	Equivalent stress in fitting state	N/mm²
$\sigma_{z,M}$	Maximum tensile stress in fitting state	N/mm²
σ_z	Maximum tensile stress in operating state	N/mm²
τ_t	Maximum torsional stress in thread	N/mm²
φ	Pitch angle	°
Φ	Force ratio	–
Φ_n	Force ratio at $n < 1$	–

Calculations for threaded fasteners
The prevailing basis for calculating high-stress threaded fasteners is VDI Directive 2230 [25]. It can be used as a simple, sufficiently precise calculation of the forces and stresses in the thread and the required tightening and loosening torques for a cylindrical single-threaded fastener, which can be considered as a section of a highly rigid multiple-threaded fastener. This means that, in many cases, even complex multiple-threaded fasteners can be considered as single-threaded fasteners. The precondition for this is that the fastener axes are parallel to each other and perpendicular to the parting planes. The components must also be elastic. Another important factor here is that only centrally preloaded and centrally stressed fasteners are considered. For large, eccentric stresses, which may cause the parting line to gape open, please refer to VDI 2230 (Figure 1).

Property classes
According to DIN EN ISO 898-1 [26], the properties of a fastener are identified by two numbers separated by a decimal point. The first number equals 1/100th of the minimum tensile strength; the second is a number that is 10 times the ratio of the yield point in relation to the tensile strength. Multiplying the two numbers gives 1/10 of the minimum yield point (example: 8.8 $\rightarrow R_e = R_{p0.2}$ = 640 MPa).

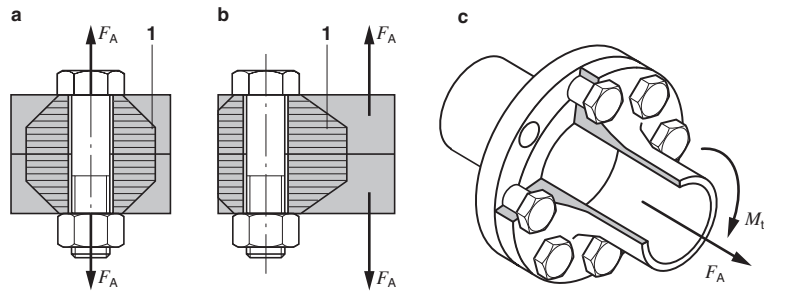
The property class of a standard nut is identified by one number. This number corresponds to 1/100 of the minimum tensile strength of a bolt of the same property class. To optimize material exploitation, therefore, bolts and nuts of equal property classes should always be paired (e.g. bolt 10.9 with nut 10).

Tightening of threaded fasteners
Prestress
Threaded fasteners are preloaded joints in which the fastener is expanded by f_{SV} by being tightened, and the parts or plates to be tightened are pressed together by f_{PV}. The deformation depends on the dimensions (cross-section and length) and on the materials (moduli of elasticity). According to Hooke's Law, in the elastic range, these are proportional to the prevailing linear force. The ratio of force F to change in length f is the spring rate

$$R = \frac{F}{f} = \frac{E \cdot A}{l}.$$

If the rigidities of fasteners and stressed parts are known (they can be calculated according to VDI 2230), the preloaded threaded fastener can be represented by a load-extension diagram. After assembly, an equilibrium of forces arises so that the preload in the bolt and the stressed parts are of identical size (Figure 2a).

Figure 1: Stresses applied to threaded fasteners
a) Centrally stressed and centrally loaded fastener,
b) Eccentrically stressed and eccentrically loaded fastener,
c) Multiple-threaded fastener.
1 Area of pressure application.



Operating forces

In the case of transversely stressed threaded fasteners (operating force F_Q perpendicular to fastener axis), the forces are transferred in the parting plane by friction. Provided that the frictional forces, which are generated by the fastener preload, are greater than the operating forces to be transferred, the assembly load-extension diagram does not change. This means that the fastener "notices" nothing about the external stress.

At a coefficient of friction μ_T in the parting line, n_S number of fasteners, i number of friction surface pairs, and S_R factor of safety against sliding, the required minimum clamping force is calculated as follows:

$$F_{K,min} = F_V \geq \frac{S_R F_Q}{\mu_T n_S i} .$$

If an external operating force F_A acts in the direction of the fastener axis according to Figure 1, the bolt is extended by f_A. At the same time the compression of the stressed parts is reduced by the same amount. The bolt is thus subjected to an additional stress of F_{SA}, whereas the stressed parts are relieved by F_{PA}. The additional fastener force F_{SA} thus depends on the rigidity and yield of the fastener (Figure 2b).

The "softer" the fastener (tensile bolt: long and thin), the smaller the additional bolt stress F_{SA} caused by an external axial operating force F_A. This must be the objective, particularly with dynamic operating forces (e.g. for cylinder-head bolts).

Force application

The rigidities of the fastener and the stressed parts also depend on force application. If the external axial operating force F_A is applied directly to the fastener head, the force-application factor $n = 1$. For the application of force in the parting line $n = 0$. Real force applications range between these two extreme values (Figure 3). In this case, only part of the stressed parts is relieved. This hardens the spring rate R_P of the stressed parts, since grip is reduced. The stressed parts are hammered onto the fastener, which then becomes apparently longer and therefore softer. Small n values therefore result in small additional fastener forces. This has a good impact on the fastener safety factor. However, the clamping force is reduced at the same time, and this is not good for the joint function.

There is no simple method of calculating the force-application factor n, which defines the point of force application. Either n ($0 \leq n \leq 1$) is estimated, or an approximate calculation is made according to VDI 2230. The forces specified in the load-extension diagram can be calculated in accordance with Table 2.

Fastener forces and torques

Calculation model

The simplest way to represent force ratios in a threaded fastener is by concentrating the surface pressure distributed to all thread turns on a single nut element. During tightening and loosening, the nut

Figure 2: Tightening threaded fasteners
a) Load-extension diagram on assembly,
b) Load-extension diagram with axial operating force F_A.

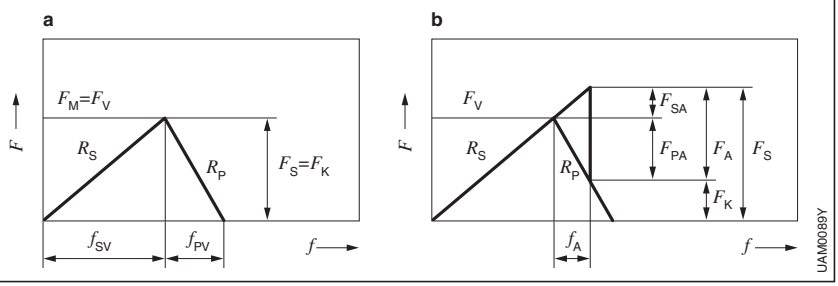


Table 2: Fastener forces (depending on force application)

Force	Force application at fastener head $n=1$	Force application any $0<n<1$	Force application in parting line $n=0$
Max. fastener force	$F_S=F_V+\Phi F_A$	$F_S=F_V+\Phi_n F_A$	$F_S=F_V$
Clamping force	$F_K=F_V-(1-\Phi)F_A$	$F_K=F_V-(1-\Phi_n)F_A$	$F_K=F_S-F_A$
Additional fastener force	$F_{SA}=\Phi F_A$	$F_{SA}=\Phi_n F_A$	$F_{SA}=0$
Additional plate force	$F_{PA}=(1-\Phi)F_A$	$F_{PA}=(1-\Phi_n)F_A$	$F_{PA}=F_A$

Where $\Phi=\frac{R_S}{(R_P+R_S)}$ and $\Phi_n=n\,\Phi$.

element moves along the bolt thread, which, if unwound, represents a slanting plane or a wedge (Figure 4).

Tightening a threaded fastener
When tightened, the nut element is pushed up the wedge by the peripheral force F_U. The resulting normal force F_N causes a frictional force F_R which acts in the opposite direction and includes the angle of friction ρ. But since all standardized thread profiles have sloping flanks, as shown in the section in Figure 4, the actual normal force F_N is perpendicular to the sloping thread surface. Therefore only the component $F'_N=F_N\cdot\cos\alpha/2$ appears in the plane force polygon. The following then applies to the frictional force:

$$F_R=F_N\cdot\mu_G=F'_N\cdot\mu'_G.$$

In order to calculate the force polygon in a plane parallel to the fastener axis, an apparent coefficient of friction is introduced:

$$\mu'_G=\frac{\mu_G}{\cos\alpha/2}=\tan\rho'.$$

When the peripheral force acts on the flank diameter d_2, the thread torque becomes

$$M_G=F_V\cdot\frac{d_2}{2}\cdot\tan(\varphi+\rho').$$

Figure 3: Force application between fastener head and parting line
n Force-application factor.

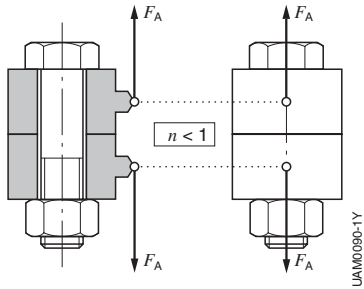
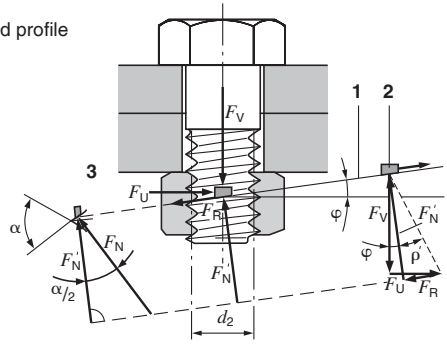


Figure 4: Forces for tightening a threaded fastener
1 Bolt thread,
2 Nut element,
3 Section through thread profile and nut element.



To tighten a bolt to preload F_V, a head friction torque M_{KR} is required to overcome the friction between the head and nut bearing surfaces, in addition to the thread torque M_G. At a coefficient of friction μ_K and a mean head friction diameter D_{Km}, the head friction torque is calculated as follows:

$$M_{KR} = F_V \cdot \mu_K \cdot \frac{D_{Km}}{2}.$$

The bolt tightening torque applied during assembly is then:

$$M_A = M_G + M_{KR} = F_V \left(\frac{d_2}{2} \tan(\varphi + \rho') + \mu_k \frac{D_{km}}{2} \right).$$

Loosening a threaded fastener
When a threaded fastener is loosened, the frictional force changes in the opposite direction to the tightening action. The bolt loosening torque (release torque) is calculated as follows:

$$M_L = F_V \left(\frac{d_2}{2} \tan(\varphi + \rho') - \mu_k \frac{D_{km}}{2} \right).$$

In the case of self-locking threads ($\varphi < \rho'$), the loosening torque becomes negative. This means that a torque must be applied in the opposite direction to the tightening action.

Design of threaded fasteners

Overstressing
If the minimum thread reach $m = (1.0 \text{ to } 1.5) \cdot d$ is maintained, a threaded fastener fails if overstressed, not because the thread turns are stripped, but because the cylindrical screw bolt ruptures.

Assembly stress
When the fastener is tightened to preload F_V, it is stressed up to tensile stress. Due to the thread torque M_G, it is also stressed to torsion. As the friction in the thread prevents the bolt from turning back, the torsional stress also acts after tightening. According to the distortion-energy theory (see Mechanics, Basic principles) the equivalent stress in the bolts is

$$\sigma_{red,M} = \sqrt{\sigma_{z,M}^2 + 3\,\tau_t^2} \le v \cdot R_{p0.2}$$

with tensile stress

$$\sigma_{z,M} = \frac{F_{V,max}}{A_S} = \frac{\alpha_A \cdot F_V}{A_S}$$

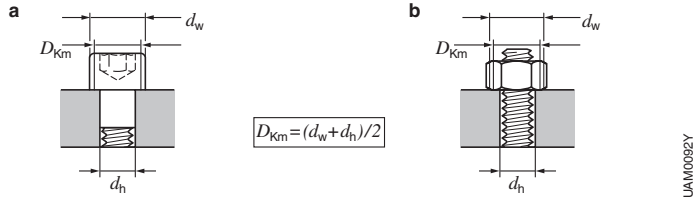
and torsional stress

$$\tau_t = \frac{M_{G,max}}{W_t} = \frac{16 \cdot \alpha_A \cdot F_V \cdot d_2 \cdot \tan(\varphi + \rho')}{2 \cdot \pi \cdot d_3^3}.$$

The tightening factor α_A takes account of the imprecision which is unavoidable during assembly. For torque-controlled tightening (torque wrench) it is $\alpha_A = 1.4$ to 1.6, for pulse-controlled tightening (impact wrench) it is $\alpha_A = 2.5$ to 4.0.

To ensure high functional security, it is necessary to aim for the highest possible material exploitation. This is taken into account by efficiency v. Table 3 shows the permissible preload forces and tightening torques during assembly for different coefficients of friction at 90 % efficiency (v 0.9) of the standardized minimum yield point.

Figure 5: Effective diameter D_{Km} for head friction torque
a) For Allen screw (d_w = head diameter),
b) For hexagonal bolt/nut (d_w = width across flats).



Static stress
An axial operating force F_A increases tensile stress in the bolt. Since, in the operating state, the effect of torsional stress is less than in the fitted state, VDI 2230 applies to reduced stress:

$$\sigma_{red,B} = \sqrt{\sigma_z^2 + 3(0.5 \cdot \tau_t)^2} < R_{p0.2}$$

at tensile stress:

$$\sigma_z = \frac{F_{S,max}}{A_S} = \frac{\alpha_A \cdot F_V + F_{SA}}{A_S}.$$

Vibrational stress
At a dynamic operating force F_A, the variable stress component σ_a must not exceed the permissible variable stress σ_A. The following applies:

$$\sigma_a = \frac{F_{S,max} - F_{S,min}}{2 \cdot A_S} \leq \sigma_A.$$

The permissible variable stress σ_A does not depend on the property class, but on the nominal diameter only (Table 4).

Surface pressure between head and nut bearing surface
In the case of a large preload, the surface pressure on the head and the nut bearing surfaces must be checked. Excessive surface pressures can cause plastic deformation and loss of preload. This can result in threaded fasteners coming loose.

The surface pressure p resulting from the maximum fastener force must not exceed the permissible surface pressure limit p_G (guideline values for p_G, see Table 5):

$$p = \frac{4 F_{S,max}}{\pi\left(d_w^2 - d_h^2\right)} \leq p_G.$$

Table 3: Permissible preloads and tightening torques for regular bolts (acc. to DIN 2230)

Thread	Assembly preload F_V for different coefficients of friction μ in thread						Tightening torque M_A with thread friction μ = 0.12					
	F_V in 10^3 · N for μ = 0.1			F_V in 10^3 · N for μ = 0.2			M_A in Nm with μ_K = 0.1			M_A in Nm with μ_K = 0.2		
	8.8	10.9	12.9	8.8	10.9	12.9	8.8	10.9	12.9	8.8	10.9	12.9
M4	4.5	6.7	7.8	3.9	5.7	6.7	2.6	3.9	4.5	4.1	6.0	7.0
M5	7.4	10.8	12.7	6.4	9.4	11.0	5.2	7.6	8.9	8.1	11.9	14.0
M6	10.4	15.3	17.9	9.0	13.2	15.5	9.0	13.2	15.4	14.1	20.7	24.2
M8	19.1	28.0	32.8	16.5	24.3	28.4	21.6	31.8	37.2	34.3	50.3	58.9
M10	30.3	44.5	52.1	26.3	38.6	45.2	43	63	73	68	100	116
M12	44.1	64.8	75.9	38.3	56.3	65.8	73	108	126	117	172	201
M14	60.6	88.9	104.1	52.6	77.2	90.4	117	172	201	187	274	321
M16	82.9	121.7	142.4	72.2	106.1	124.1	180	264	309	291	428	501
M18	104	149	174	91.0	129	151	259	369	432	415	592	692
M20	134	190	223	116	166	194	363	517	605	588	838	980
M22	166	237	277	145	207	242	495	704	824	808	1151	1347
M24	192	274	320	168	239	279	625	890	1041	1011	1440	1685
M27	252	359	420	220	314	367	915	1304	1526	1498	2134	2497
M30	307	437	511	268	382	447	1246	1775	2077	2931	2893	3386

Threaded-fastener locking devices
The spontaneous loosening of a threaded fastener is caused by complete or partial loss of preload, which in turn is caused by settling events (loosening) or relative movements in the parting line (unscrewing).

Loosening
Contact stress f_z caused by plastic deformation results in loss of preload

$$F_z = \frac{R_p \cdot R_S}{R_p + R_S} \cdot f_z \,.$$

Contact stress f_z in µm depends on the surface properties and the number of parting lines (Table 6).

Total contact stress equals the sum of individual component parts. However, contact stress determined in this way is only valid if the surface pressure limits are not exceeded. Otherwise, significantly greater settling will arise. Locking devices are intended to reduce or compensate for settling.

The following measures provide reliable protection against loosening:
– High preload,
– Elastic threaded fasteners,
– Low surface pressures due to large bearing surfaces and sufficient thread reach,
– Low number of parting lines,
– No plastic or quasi-elastic elements (e.g. seals) with stress.

Unscrewing
Dynamic stresses, particularly perpendicular to the fastener axis, may cause threaded fasteners to come loose despite sufficient preload. If transverse movements can occur, locking devices to prevent unscrewing insure that the joint retains its function. Suitable measures include:
– Avoid transverse movements by positive locking in the parting line,
– Elastic threaded fasteners
– Large grip lengths,
– High preload,
– Suitable locking devices (locking or bonding elements, Figure 6).

Table 4: Permissible variable stress σ_A

Diameter range	M6 to M8	M10 to M18	M20 to M30
Permissible variable stress σ_A in N/mm²	60	50	40

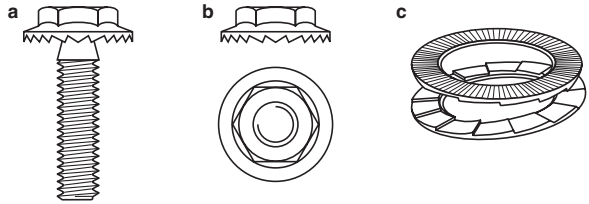
Table 5: Standard values for surface pressure limit p_G as per VDI 2230

Material	Surface pressure limit p_G in N/mm²
GD-AlSi 9 Cu 3	290
S 235 J	490
E 295	710
EN-GJL-250	850
34 CrNiMo 6	1,080

Table 6: Contact stress f_z depending on surface properties and number of parting lines

Surface	Load	Contact stress f_z in µm		
		in thread	per head or nut bearing surface	per parting line
$R_z < 10$	Tension/Compression Shear	3.0 3.0	2.5 3.0	1.5 2.0
$10 \le R_z < 40$	Tension/Compression Shear	3.0 3.0	3.0 4.5	2.0 2.5
$40 \le R_z < 160$	Tension/Compression Shear	3.0 3.0	4.0 6.5	3.0 3.5

Figure 6: Locking devices (examples)
a) Flat-headed self-locking screw, b) Flat-headed self-locking nut, c) Locking washer pair.



UAM0113Y

Thread dimensions

Figure 7: Metric ISO thread
(DIN 13, ISO 965); nominal dimensions.

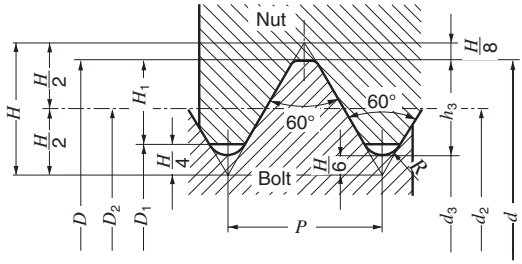


Table 7: Metric standard thread
Designation example: M8 (nominal thread diameter 8 mm).

Nominal thread dia.	Pitch	Pitch diameter	Minor diameter		Thread depth		Stress area
$d=D$	P	$d_2=D_2$	d_3	D_1	h_3	H_1	A_s in mm²
3	0.5	2.675	2.387	2.459	0.307	0.271	5.03
4	0.7	3.545	3.141	3.242	0.429	0.379	8.78
5	0.8	4.480	4.019	4.134	0.491	0.433	14.2
6	1	5.350	4.773	4.917	0.613	0.541	20.1
8	1.25	7.188	6.466	6.647	0.767	0.677	36.6
10	1.5	9.026	8.160	8.376	0.920	0.812	58.0
12	1.75	10.863	9.853	10.106	1.074	0.947	84.3
14	2	12.701	11.546	11.835	1.227	1.083	115
16	2	14.701	13.546	13.835	1.227	1.083	157
20	2.5	18.376	16.933	17.294	1.534	1.353	245
24	3	22.051	20.319	20.752	1.840	1.624	353

Table 8: Metric fine thread
Designation example: M8 x 1 (nominal thread diameter 8 mm and pitch 1 mm).

Nominal thread dia.	Pitch	Pitch diameter	Minor diameter		Thread depth		Stress area
$d=D$	P	$d_2=D_2$	d_3	D_1	h_3	H_1	A_s in mm²
8	1	7.350	6.773	6.917	0.613	0.541	39.2
10	1.25	9.188	8.466	8.647	0.767	0.677	61.2
10	1	9.350	8.773	8.917	0.613	0.541	64.5
12	1.5	11.026	10.160	10.376	0.920	0.812	88.1
12	1.25	11.188	10.466	10.647	0.767	0.677	92.1
16	1.5	15.026	14.160	14.376	0.920	0.812	167
18	1.5	17.026	16.160	16.376	0.920	0.812	216
20	2	18.701	17.546	17.835	1.227	1.083	258
20	1.5	19.026	18.160	18.376	0.920	0.812	272
22	1.5	21.026	20.160	20.376	0.920	0.812	333
24	2	22.701	21.546	21.835	1.227	1.083	384
24	1.5	23.026	22.160	22.376	0.920	0.812	401

Figure 8: Pipe threads for non self-sealing joints
(DIN ISO 228-1); parallel internal threads and external threads; nominal dimensions

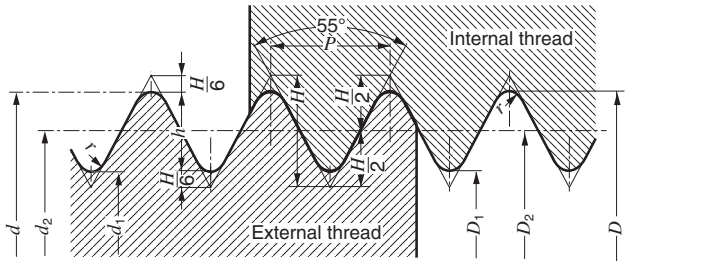


Table 9: Designation example: G1/2 (nominal thread size 1/2)

Nominal thread size	Number of threads per inch	Pitch P mm	Thread depth h mm	Outside diameter $d=D$ mm	Pitch diameter $d_2=D_2$ mm	Minor diameter $d_1=D_1$ mm
1/4	19	1.337	0.856	13.157	12.301	11.445
3/8	19	1.337	0.856	16.662	15.806	14.950
1/2	14	1.814	1.162	20.955	19.793	18.631
3/4	14	1.814	1.162	26.441	25.279	24.117
1	11	2.309	1.479	33.249	31.770	30.291

Figure 9: Whitworth pipe threads for threaded pipes and fittings
(DIN 10226-1); parallel internal threads and tapered external threads; nominal dimensions (mm).

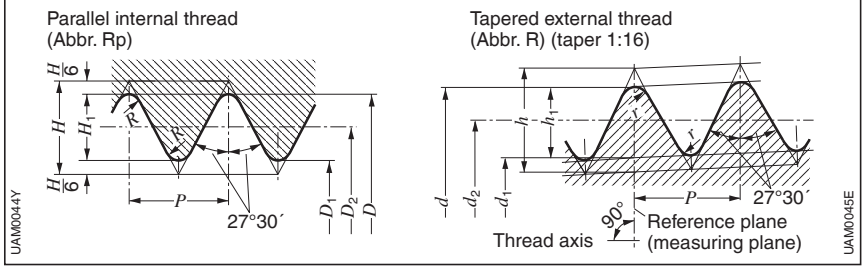


Table 10: Whitworth pipe threads for threaded pipes and fittings

Code		Outside diameter	Pitch diameter	Minor diameter	Pitch	Number of threads per inch
External thread	Internal thread	$d=D$	$d_2=D_2$	$d_1=D_1$	P	Z
R 1/4	Rp 1/4	13.157	12.301	11.445	1.337	19
R 3/8	Rp 3/8	16.662	15.806	14.950	1.337	19
R 1/2	Rp 1/2	20.955	19.793	18.631	1.814	14
R 3/4	Rp 3/4	26.441	25.279	24.117	1.814	14
R 1	Rp 1	33.249	31.770	30.291	2.309	11

Areas of application: For joining parallel internal threads to valves and fittings, threaded flanges, etc. with tapered external threads.

Snap-on connections on plastic components

Features

Snap-on connections are a cheap and efficient way of fitting plastic components. They are used to connect housing halves, on plug connectors, and to secure mounting parts in plastic housings. They exploit the high expansibility of the plastics with relatively low rigidity.

All snap-on connections are characterized by the brief excursion of a resilient element in the joining process before it snaps into place behind a locating lug. Depending on the configuration of the joining angles at the snap-on elements, it is possible to produce nondestructively detachable and nondetachable connections (Figure 1).

The basic shapes [27] of snap-on connections are (Table 1):

– Resilient snap-on hooks (bending springs fixed on one side),
– Torsion snap-on hooks,
– Resilient clips,
– Ring-shaped snap-on connections, also segmented (slotted lengthways),
– Spherical snap-on connections, also segmented.

Figure 1: Snap-on connection (principle)
a) Decisive variables,
b) Joining and release angles (detachable connection: $\alpha_2 < 90°$, nondetachable connection: $\alpha_2 \geq 90°$).
1 Spring element, 2 Locating lug.
f Spring travel (rear section), l Length, h Thickness at fixation cross-section, F Joining force, Q Excursion force, α_1 Joining angle, α_2 Release angle.

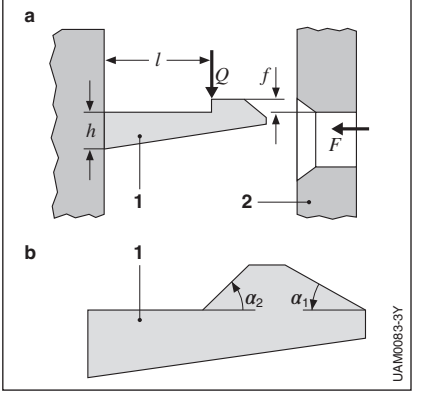
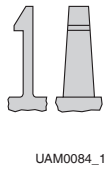
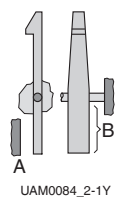
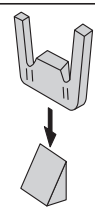
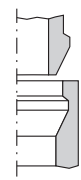
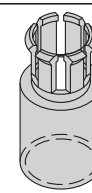


Table 1: Basic shapes and types of snap-on connection [27]

Shape	Hook shape			Ring shape		
				Annular ring, ring groove	Annular ring segmented, ring groove	Hollow-sphere section
Spring element	Bending spring	Torsion spring (+ bending spring)	Connected bending springs with locking element	Annular spring	Annular spring, segmented	Annular spring
Designation	(Bending) snap-on hook	Torsion snap-on hook	Resilient clip	Ring snap-on element	Ring snap-on connection	Spherical snap-on element
Type	 UAM0084_1Y	 UAM0084_2-1Y	 UAM0084_3Y	 UAM0084_4Y	 UAM0084_5Y	 UAM0084_6Y

A Stop against overextension of the torsion axis.
B Button side for releasing the torsion snap-on hook must be more stiffly designed than the side with engagement hook.

Design guidelines and layout

The spring elements are designed to accommodate the permitted elongation of the plastic in the joining process. The least favorable material condition must be taken into account here (e.g. dry polyamide).

The elongation-dependent secant modulus

$$E_s = \frac{\sigma_1}{\varepsilon_1}$$

is used (determination depicted in Figure 2). The corresponding stress-strain diagrams can be taken for example from the CAMPUS database [29] or obtained from the material manufacturers.

In order to achieve a uniform distribution of strain and optimum material utilization in the bending range of the spring elements, the thickness from the root to the free end should decrease by half. As an alternative, the width to the end of the hook can be reduced to one quarter. Radii at the point of connection of the spring element to the component are recommended in that they help to eliminate concentrations of strain.

When joined, the spring element must be fully returned to its initial state in order to prevent creeping under load and thus permanent deformation. Tensile stress in the snap-on element as a result of operational forces is permitted.

The permitted excursion (spring travel f) in joining is dependent on the geometry of the snap-on hook and the permitted elongation ε of the plastic (Table 2). Formulas for different cross-sectional shapes can be taken from the relevant technical literature ([28], [30], [31]) or are provided by special calculation programs.

Included in the calculation of the excursion force Q are the rigidity of the plastic as the secant modulus E_s and the geometry as the bending moment/section modulus W.

The joining force F is calculated from the excursion force Q, the joining angle α_1 (usually 30°), and the coefficient of friction μ between the joining components according to the formula:

$$F = \frac{Q\,(\mu + \tan\alpha_1)}{(1 - \mu\,\tan\alpha_1)}\,.$$

Values for the section modulus W and the friction coefficient μ can also be taken from the technical literature ([28], [30], [31]).

The release force of a snap-on connection is calculated according to the same formula as the joining forced, where the release angle α_2 of the snap-on hook (usually 60°) is to be used. In the case of a nondetachable connection ($\alpha_2 \geq 90°$), the thrust capacity of the snap-on element limits the strength.

Figure 2: Determination of secant modulus E_s
$E_{S1} = \sigma_1/\varepsilon_1$,
E_0 Modulus of elasticity.

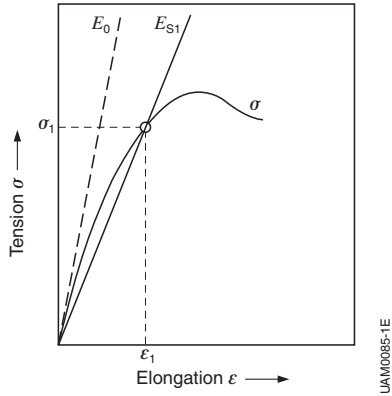


Table 2: Reference values for permitted elongation ε for snap-on connections. [28]
(For single, brief joining operation, in case of frequent joining operations reduction of the values by approx. 40 %).

	Material	ε
semi-crystalline	PE	0.080
	PP	0.060
	PA (conditioned), POM	0.060
	PA (dry)	0.040
	PBT	0.050
amorphous	PC	0.040
	ABS	0.025
	PVC	0.020
	SAN	0.020
	PS	0.018
glass-fiber-reinforced	PA-GF30 (conditioned)	0.020
	PA-GF30 (dry)	0.015
	PA-GF50 (dry)	0.005
	PBT-GF30	0.015
	PC-GF30	0.018
	ABS-GF30	0.012

Calculation programs
Different plastic manufacturers provide easy-to-use calculation programs as a customer service (e.g. "Snaps" from BASF [32], "FEMsnap tool" from Covestro [33], "FEMSnap" from Lanxess [34]). Most of the material data for manufacturers' product ranges are integrated into these programs. It is also possible and recommended to check the snap-on-hook design by means of FEM analysis within the framework of component development.

References

[1] H. Haberhauer; F. Bodenstein: Maschinenelemente. 18th Edition, Verlag Springer Vieweg, 2017.
[2] Dubbel, Taschenbuch für den Maschinenbau. 24th Ed., Verlag Springer Vieweg, 2014.
[3] F.G. Kollmann: Welle-Nabe-Verbindungen, Springer-Verlag, 1984.
[4] DIN 471: Circlips (retaining rings) for shafts – Normal type and heavy type.
[5] DIN 981: Rolling bearing – Locknuts.
[6] DIN 6880: Bright key steel – dimensions, permissible variations, weights.
[7] DIN 6885-1: Drive type fastenings without taper action; parallel keys, keyways, deep pattern.
[8] ISO 14: Straight-sided splines for cylindrical shafts with internal centering; dimensions, tolerances and verification.
[9] DIN 5464: Straight-sided splines – Heavy series.
[10] DIN 5481: Serration splines.
[11] DIN 5480: Involute splines based on reference diameters.
[12] DIN 5482-1: Internal and external involute spline profiles; dimensions.
[13] DIN 7190: Interference fits – Calculation and design rules.
[14] DIN 254: Geometrical product specifications (GPS) – Series of conical tapers and taper angles; values for setting taper angles and setting heights.
[15] DIN 406-10: Engineering drawing practice; dimensioning; concepts and general principles.
[16] DIN 3990-1: Calculation of load capacity of cylindrical gears; introduction and general influence factors.
[17] DIN-Taschenbuch 10: Mechanische Verbindungselemente 1 – Schrauben, Nationale Normen. 24th Ed., Beuth-Verlag, 2017.
[18] DIN-Taschenbuch 45: Gewinde. 11th Ed., Beuth-Verlag, 2017.
[19] DIN-Taschenbuch 140: Mechanische Verbindungselemente 4 – Muttern. 10th Ed., Beuth-Verlag, 2016.
[20] H. Wiegand; K.-H. Kloos; W. Thomala: Schraubenverbindungen – Grundlagen, Berechnung, Eigenschaften, Handhabung. 5th Edition, Springer-Verlag, 2007.
[21]: DIN 13: ISO general purpose metric screw threads.
[22]: DIN 965: ISO general purpose metric screw threads.
[23] DIN EN ISO 228: Pipe threads where pressure-tight joints are not made on the threads.
[24] DIN EN 10226-1: Pipe threads where pressure-tight joints are made on the threads.
Part 1: Taper external threads and parallel internal threads; dimensions, tolerances and designation.
[25] VDI Directive 2230: Systematische Berechnung hochbeanspruchter Schraubenverbindungen (Systematic calculation of high-stress threaded fasteners). Beuth-Verlag: Sheet 1, 2015; Sheet 2, 2014.
[26] DIN EN ISO 898-1: Mechanical properties of fasteners made of carbon steel and alloyed steel. Part 1: Bolts, screws and studs with specified property classes – Coarse thread and fine pitch thread.
[27] U. Delpy u.a.: Schnappverbindungen aus Kunststoff. Expert Verlag, 1989.
[28] A. Maszewski.: Schnappverbindungen und Federelemente aus Kunststoff. Anwendungstechnische Information, Bayer AG, 2000.
[29] http://www.campusplastics.com.
[30] G.W. Ehrenstein (Editor): Handbuch Kunststoffverbindungstechnik. Carl Hanser Verlag, 2004.
[31] T. Brinkmann: Produktentwicklung mit Kunststoffen. Carl Hanser Verlag, 2010.
[32] http://www.plasticsportal.net/wa/plasticsEU~de_DE/portal/show/common/content/technical_resources/calculation_programmes.
[33] http://www.plastics.covestro.com/en/Engineering/Tools/FEMSnap-tool.
[34] https://techcenter.lanxess.com/scp/emea/en/techServscp/article.jsp?docId.

Permanent joints

Welding

Automotive components and subassemblies are joined using a wide and highly varied range of welding and bonding techniques [1], [2]. Resistance and fusion welding are among the most commonly applied welding methods. The overview in Figure 1 shows the most important resistance-welding procedures used in production technology (for procedure types and symbols, see DIN 1910, Part 100, [3]).

Resistance welding

Resistance spot welding

In resistance spot welding, locally applied electrical current is used to melt the contact surfaces of the parts to be joined into a soft or fluid state. The parts are then joined together under pressure (Figure 2a and 2b). The spot-welding electrodes which conduct the welding current also convey the electrode force to the parts to be joined (joining parts). The amount of heat required to create the welding spot is determined in accordance with the equation

$Q = I^2 R t$ (Joule's Law).

The precise amount of heat required Q is a function of welding-current intensity I, resistance R, and welding time t. A good and adequate spot diameter can be achieved by the coordination of welding-current intensity I, electrode force F, and welding time t.

According to the manner in which the current is conducted, a distinction is drawn between bilateral direct resistance spot welding (Figure 2a) and unilateral indirect resistance spot welding (Figure 2b).

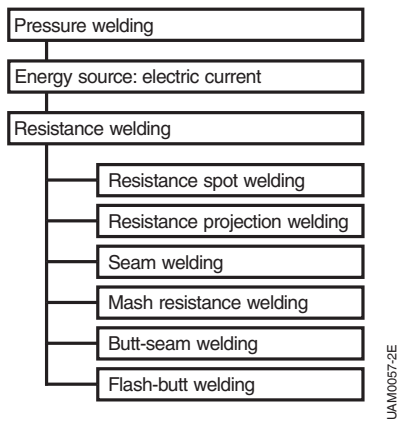
The electrode for a specific spot welding operation is selected with reference to shape, outside diameter and point diameter. Since the joining parts should be as free as possible of scaling, oxides, paints, grease, and oils, they receive where necessary appropriate surface pretreatment prior to welding.

Applications:
- Joining of sheet parts up to an approximate 3 mm single sheet thickness in the lap joint or as a welding flange.
- Multiple-sheet as well as two-sheet joints of different sheet thicknesses and sheet materials.
- Spot-weld bonding in combination with adhesive.

Resistance projection welding

Projection welding (Figure 2c) is a process in which electrodes with a large surface area are employed to conduct the welding current and the electrode force to the workpiece. The projections, which are generally incorporated into the thicker of the joining parts, cause the current to concentrate at the contact points. The electrode force levels the projections partially or completely again during the welding process. A permanent, inseparable joint is produced at the welding seams, i.e. the contact points. One or more than one projection can be welded simultaneously, depending on the type of projection (round, longitudinal or annular) and the power available from the welding appa-

Figure 1: Classification of resistance-welding process in acc. with DIN 1910-100



ratus. Depending on the number of spots welded, a distinction is drawn between single-projection welding and multiple-projection welding.

Resistance projection welding requires high welding currents applied for short periods of time.

Applications:
– Welding parts of different thicknesses.
– Welding multiple projections in a single operation.

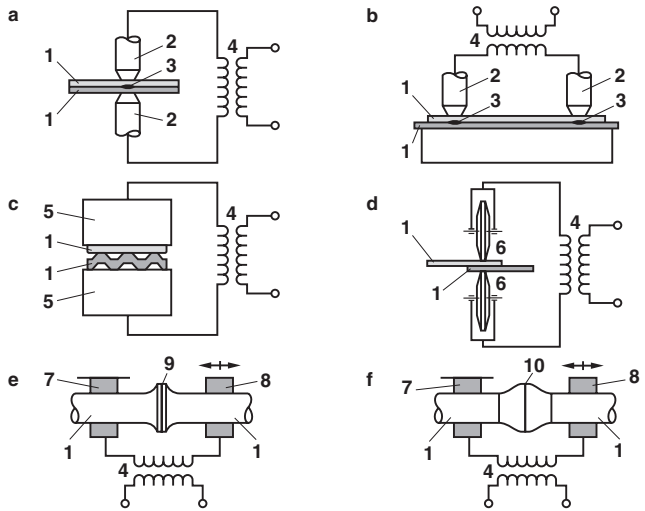
Seam welding
In this process, roller electrodes replace the spot-welding electrodes used in resistance spot welding (Figure 2d). Contact between the roller set and the workpiece is limited to an extremely small surface area. The roller electrodes conduct the welding current and apply the electrode force. Their rotation is coordinated with the movement of the part.

Applications:
– Production of sealed welds or seam spot welds (e.g. fuel tanks).

Flash-butt welding
In flash-butt welding, the butt ends of the workpieces are joined under light pressure (Figure 2e) while the flow of current at the contact surfaces produces localized heat and melting (high current density) (current supply via copper jaws). The metal's vapor pressure drives molten material from the contact patches (burn-off). The welded joint is then created by upsetting the workpieces with a high compressive force.

The butt ends should be parallel to each other and at right angles to the direction in which the force is applied (or virtually so). Smooth surfaces are not required. A certain amount of extra length must be factored in to compensate for the losses incurred in the flash-butt welding process.

Figure 2: Resistance-welding processes
a) Bilateral resistance spot welding, b) Unilateral resistance spot welding,
c) Resistance projection welding, d) Seam welding, e) Flash-butt welding, f) Butt-seam welding.
1 Part to be joined, 2 Spot-welding electrodes, 3 Welding spot, 4 Transformer,
5 Large-surface electrodes, 6 Roller electrodes, 7 Fixed copper jaws,
8 Longitudinally movable copper jaws, 9 Weld joint with seam, 10 Weld joint with bead.



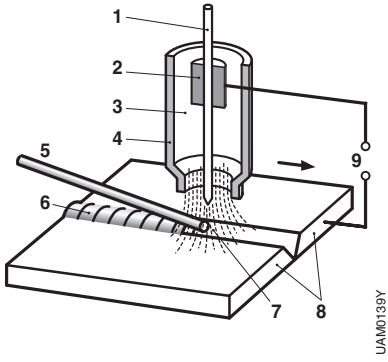
The result is a weld with the characteristic projecting seam.

Applications:
– Joining in the butt joint, e.g. rims, link chains.
– Workshop procedures, e.g. for saw bands.

Butt-seam welding
This process (Figure 2f) employs copper jaws to conduct the welding current to the workpieces to be joined. When the welding temperature is reached, the current switches off. Constant pressure is maintained and the workpieces then weld together (requirement: properly machined butting faces). The process does not completely displace contamination that may be present at the butt ends. The result is a weld with the characteristic projecting bead.

Applications:
– Joining in the butt joint, e.g. shafts, axles.

Figure 3: Principle of tungsten inert-gas welding
1 Tungsten electrode,
2 Current contact tube,
3 Inert gas,
4 Inert-gas nozzle,
5 Welding filler,
6 Weld seam,
7 Arc,
8 Workpieces,
9 Energy source.



Fusion welding
The term "fusion welding" describes a process employing limited local application of heat to melt and join the parts. No pressure is applied. Shielded (inert gas) arc welding is a type of fusion welding. The electrical arc extending between the electrode and the workpiece serves as the heat source. Meanwhile, a layer of inert gas shields the arc and the melted area from the atmosphere. The type of electrode is the factor which distinguishes between various techniques:

Tungsten inert-gas welding
In this process, an arc is maintained between the workpiece and a stable, non-melting tungsten electrode. The shielding (inert) gas is argon or helium. The rod-shaped weld metal is supplied from the side (Figure 3).

Gas-shielded metal-arc welding
In gas-shielded metal-arc welding, an arc is maintained between the melting end of the wire electrode (material feed) and the workpiece. The welding current flows to the wire electrode via a current-contact nozzle in the welding torch. Metal inert-gas (MIG) welding uses inert gases (slowly reacting and noble gases such as argon, helium, or combinations of the two) as protective gas. MIG welding is used with materials which are particularly sensitive to oxidation, e.g., aluminum, magnesium, titanium, and nickel alloys.

Metal active-gas (MAG) welding, on the other hand, uses an active gas (e.g., CO2 or mixed gases containing CO2, argon and sometimes oxygen). MAG welding is used above all for unalloyed and low-alloy steels. Using inert gases with low active-gas admixtures for high-alloy, e.g., stainless steels is also referred to as MAG welding.

Laser-beam welding
Laser-beam welding (or laser welding for short) involves the use of light as the energy source to melt on the workpieces to be welded. The monochromatic laser radiation is generated in a laser source, where the wavelength is determined by the respective excitation medium. CO_2 lasers and NdYAG lasers (solid-state lasers) are used in industrial practice [5], [6]. Newer developments are diode lasers and fiber lasers.

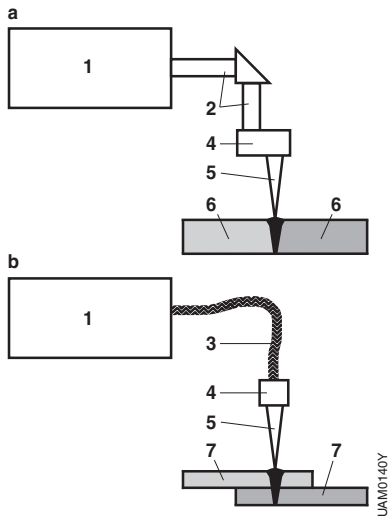
Depending on the wavelength, tube guides with refraction mirrors (CO_2 lasers) or optical fibers (e.g. NdYAG lasers) are required to direct the beam from the beam generator to the weld (Figure 4). To utilize the energy of the laser beam for welding, it is necessary to implement focusing with mirror or lens systems. This allows particularly high energy flux densities to be achieved at the weld, resulting in the deep-welding effect, which provides for particularly deep, but at the same time, narrow seams. In the simplest case welding is performed without filler metal.

Feeding the workpiece relative to the focusing optics, the optics relative to the workpiece or a combination of the two is used to direct the beam along the seam. Systems with optical fibers are particularly well suited to moving the focusing optics, e.g. during robot-controlled welding of three-dimensional seam geometries.

In the case of remote welding the beam is directed over the workpiece from a relatively large distance (long focal length of the focusing optics) by moving the focusing mirror or the focusing lens inside the focusing optics.

It is possible to operate several machining stations with one beam generator by using beam switches.

Figure 4: Principle of laser-beam welding
a) Beam direction by mirror,
b) Beam direction by optical fiber.
1 Beam generator,
2 Beam direction by refraction mirror,
3 Beam direction by optical fiber,
4 Focusing optics,
5 Focused laser beam,
6 Workpieces (butt joint) with weld seam,
7 Workpieces (lap joint) with weld seam.



Applications:
– Joining of steel sheets using a lap joint in body manufacturing,
– Joining of unalloyed and low-alloy steels as butt joint in chassis and assembly components,
– Joining of high-alloy steels in exhaust-gas systems,
– Joints in seat systems,
– Joining of aluminum alloys (with filler metal).

Other welding processes
The following welding processes are also used in the automotive industry [5], [6]:
– Electron-beam welding,
– Friction welding,
– Arc pressure welding (stud welding),
– Stored-energy welding (pulsed-current arc welding).

Soldering

In soldering, a supplementary material (solder) is melted onto two or more parts of similar or varying metallic composition in order to produce a permanent connection between them. Flux or protective gas is also used [7], [8], [9].

Fluxes (non-metallic substances) are applied to remove oxide layers from the surfaces of the soldering points after cleaning and to prevent new layers from forming: This makes it possible to apply a consistent coat of solder to the joint surfaces. Information on fluxes can be found in DIN EN ISO 9454-1 [10] and DIN EN 1045 [11].

The melting temperature of the solder is below that of the parts being joined. The solder is distributed along the join to produce the connection without the parts themselves being melted.

The strength of a soldered joint can be equal to that of the base material itself. To achieve the required strength, it is necessary to have narrow soldering gaps in which the adjacent, higher-strength base material prevents the solder from deforming.

Information on the temperature range, the heat source and the geometry of the joint shape (structure) is used to describe the soldered joints. In terms of the working temperature, a distinction is made between soft soldering and hard soldering. The working temperature is defined as the lowest surface temperature at the connection between the workpieces to be joined at which the solder can be melted and distributed to form a bond. Information on solders can be found in the DIN sheets DIN EN ISO 9453 [12], DIN EN ISO 12224-1 [13], DIN EN ISO 17672 [14].

In terms of the joint shapes, a distinction is basically made between close-joint soldering (soldering gap up to 0.5 mm) and V-Joint brazing (soldering gap larger than 0.5 mm).

Distinction by working temperature

Soft soldering

Soft soldering is employed to form permanent solder joints at melting temperatures below 450 °C (as with soldering tin). Soft solders which melt at temperatures of 200 °C and below are also known as quick solders.

Soft-soldering applications:

– Used in electrical engineering for contacts (e.g., flow soldering of printed-board assemblies).

Hard soldering

Hard soldering (brazing) is used to form permanent joins at melting temperatures above 450 °C (as with copper, copper/zinc and silver alloys, e.g. silver brazing filler).

Brazing applications:

– V-joint brazing: steel sheets in body manufacturing, also of different qualities and in the case of large wall-thickness differences.
– Close-joint soldering: radiators, pipes in the assembly area.

Manufacturing methods

The method used for heating provides yet another criterion for describing soldered joints. The most important types are furnace, induction, open-flame, and iron soldering.

Sweating

Heating is performed in through-type or vacuum furnaces with defined temperature and time characteristics. Before entering the furnace, the components are secured and the solder is inserted or applied as a paste. Additional fluxes are not required. The furnace atmosphere and the vacuum assume the flux function without leaving residues on the components.

Induction soldering

Heating is locally limited by inductive means.

Induction soldering is mainly used for round components, e.g., pipes for cooling water or oil.

Flame soldering
Heating is performed by individual torches or in a gas-heated system. Depending on the specific soldering operation, oxyacetylene burners (familiar from gas welding), propane torches or soldering lamps are employed. The solder is usually supplied from the side in the shape of a rod. The flux is applied separately or integrated in the solder.

Flame soldering is used for very small production batches and prototypes or to rework and repair, for example, pipes.

Iron soldering
A hand-held or mechanically guided soldering iron provides the heat. Irons can also be used to solder pretinned surfaces.

Iron soldering only used in soft soldering.

Further processes
– Salt-bath soldering,
– Dip soldering,
– Resistance soldering, as well as,
– MIG soldering,
– Plasma soldering, and
– Laser soldering.

Close-joint soldering and V-joint brazing
Furthermore, with regard to the geometry of the solder and the soldering flux, a distinction is made between close-joint soldering and V-joint brazing.

Close-joint soldering
The soldering gap is filled by the capillary effect of the liquid solder. The strength of the soldered joint is determined entirely by the strength of the base materials. Typical manufacturing processes are sweating, flame soldering and induction soldering.

V-joint brazing
The soldering gap is filled by the force of gravity. The strength of the soldered joint is determined above all by the strength of the solder filler. Typical manufacturing processes are arc and laser soldering.

Adhesive technologies

Adhesives
An adhesive is a non-metallic material which can join workpieces (parts to be joined) by means of adhesion and its internal strength (cohesion) without the structure of the part to be joined being significantly altered. An adhesive enables the creation of permanent material joints between homogeneous or heterogeneous material pairings.

Organic and inorganic adhesives as they harden build up firstly adhesion with the material surfaces by way of chemico-physical interactions (adhesion or adhesive forces) and secondly the necessary structural strength (cohesion). Depending on their chemistry, these adhesives harden for example at room temperature, at increased temperatures or under UV radiation with different mechanisms.

Hardening (cross-linking) and adhesion occur, in accordance with the chemical basic structure and formulation, over a certain period of time, depending on for example temperature or air humidity. The reaction types are called polymerization, polyaddition and polycondensation. Spatially interlinked macromolecules are the result. Adhesives can, depending on the required hardening temperature, be subdivided into the category of those that harden at room temperature or at higher temperatures (100 to 200 °C).

Two different types of adhesive are used, based on their recipe and form of presentation: single-component and two-component adhesives.

Two-component adhesives
These are adhesives which are made up of two components which must be mixed together to harden. The stoichiometric mixture ratio of the two components must be maintained exactly. Component A contains the basic resin. The second component (component B) contains a hardener which initiates the cross-linking reaction and links the resin molecules from component A to each other. An accelerator may be added to the hardener. Hardening usually occurs at room tem-

perature, but can by accelerated by the additional influence of slightly increased temperatures (80 to 120 °C) and thus completed more quickly.

Single-component adhesives
Single-component adhesives contain all the constituents necessary to form a bond in one phase. These systems contain inhibitors which prevent a premature chemical reaction (hardening) between the reacting agents present in the one phase (monomers, resin and hardener; premixed two-component adhesive). Single-component adhesives do not require the additional process step of mixing the components as a two-component adhesive requires.

Hardening must, depending on the chemical formulation, be initiated by higher temperatures (in an oven, by induction or infrared radiation), ultraviolet radiation or air humidity, and completed. In this process the inhibitors are neutralized and the accelerators also contained in the adhesive are released to speed up the hardening process.

Most single-component adhesives must be stored under cooled conditions (electrical conductive adhesives down to −20 °C in an upright deep-freezer, structural adhesives and sealants at 4 to 10 °C in a cold warehouse) or in a dark environment (UV adhesives) until they are ready for use.

Structural designs of adhesive connections
Adhesive connections should be designed so that predominantly shear loads occur. Overlapping joints are particularly suitable. Butt joints subjected to tensile or sliding forces should be avoided.

Combinations of adhesive bonding with other joining methods, e.g. welding, screwing or riveting, can have beneficial effects. The joining spots secure the components during the adhesive hardening time. Stress concentrations for example at the edges of weld spots or at rivet spots can also be minimized. Constructions of this nature provide enhanced structural integrity, rigidity, and damping when subjected to dynamic loads.

Examples of adhesives with technical application potential
The most important adhesives are epoxy resins, silicones, polyurethanes, and acrylates.

Silicone and polyurethane adhesives
Silicone and polyurethane adhesives are important particularly in bonds with dynamic loads, but also where tightness against liquid media (e.g. water) is called for. Depending on the medium (polar, nonpolar), there is however a risk of flexible adhesives swelling (e.g. in fuel). This process can be reversed provided the system can be re-dried during operation and the adhesive has not suffered any damage (e.g. material embrittlement by extraction). Thanks to their flexibility (elastic properties), these adhesives can compensate the different coefficients of thermal expansion of the parts to be joined in the temperature-application range by means of their deformability (low modulus of elasticity above a crystallization temperature of −60 °C to −40 °C) such that only low mechanical stresses can be built up in the bond. Compared with polyurethanes, silicones are also highly suitable for use at high application temperatures up to approximately 220 °C.

Epoxy adhesives
Epoxy adhesives, on the other hand, are predominantly very brittle and hard adhesives which can, depending on their formulation, be used up to approximately 200 °C. Compared with flexible adhesives, they are much less susceptible to swelling in liquid media. High-strength epoxy resins lack the ability of silicones to compensate by means of deformability (elongation) mechanical stresses in the adhesive bond which are caused by different coefficients of thermal expansion of the parts to be joined in the temperature-application range. Exceptions are adhesives which, thanks tor their special formulation (e.g. impact-resistant modification), are used specifically for bonds in body manufacturing in order specifically to reduce the mechanical forces and energies that occur in a crash situation and to maintain the structural integrity of the passenger cell. Epoxy adhesives facilitate bonds with

very high adhesive strengths (high modulus of elasticity up to the glass-transition temperature; this is in the range of 80 to 200 °C, depending on the chemical recipe of the adhesive components).

Acrylate adhesives
Acrylates are capable of fast hardening reactions and therefore highly attractive for use in production processes. These adhesives can harden, depending on their chemical architecture, in response to ultraviolet radiation, mixing of components, at increased temperatures or also in response to air humidity ("superglue") within a few minutes down to a matter of seconds. On the other hand, acrylates do not demonstrate as pronounced thermomechanical performance as silicone and epoxy adhesives above approximately 120 °C or in aggressive media.

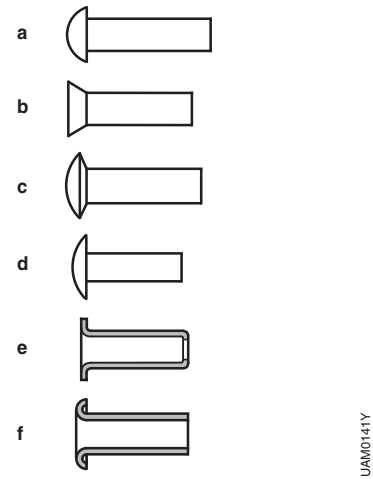
Automotive applications
Adhesive joining has become a standard technique in automotive engineering. The individual areas of application can be classified as follows:
– Electronic components: sealing/bonding of housings for sensors, ECUs and video systems; electrically and thermally conductive bonding of electronic component in power modules and electronic circuits.
– Electric motors: magnetic bonds in rotors and stators.
– Body shell: raised-seam and brace bonding for attached components.
– Assembly line: attachment of insulating material, appliqués, moldings, mirror-support bracket to windshield.
– Component production: bonding of brake pads, laminated safety glass (LSG), rubber-metal connections to absorb vibration.

Riveting

Classic riveting
Process
Riveting is used to produce a permanent fixed connection between two or more components made of identical or dissimilar materials. Here the components to be joined are pierced together by means of drilling or punching. The rivet is then inserted into the hole as the joining element. Depending on the method and application, riveted joints are divided into the following categories:
– Permanent rigid connections (interference fits, for example in mechanical and plant engineering),
– Permanent, sealed connections (for example in boilers and pressure vessels),
– Extremely tight seals (for example in pipes, vacuum equipment).

A distinction is drawn between cold and hot riveting, depending on the temperature used. Cold riveting is employed for rivet joints up to 10 mm in diameter in steel, copper, copper alloys, aluminum,

Figure 5: Rivet shapes
a) Mushroom-head rivet,
b) Countersunk-head rivet,
c) Oval countersunk-head rivet,
d) Flat round-head rivet,
e) Hollow rivet,
f) Tubular rivet.



etc. Rivets with diameters in excess of 10 mm are installed hot.

The most common rivet shapes are (Figure 5, [15]...[20]) the mushroom-head rivet (DIN 660), the countersunk-head rivet (DIN 661), the oval countersunk-head rivet (DIN 662), the flat round-head rivet (DIN 674), the hollow rivet (DIN 7339), and the tubular rivet (DIN 7340).

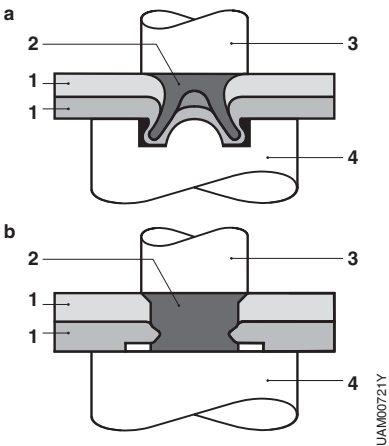
There are also standardized rivets for specialized applications, e.g. explosive rivets or blind rivets. Blind rivets are hollow and are expanded by a drift or punch.

Furthermore, rivets are frequently designed for use as function elements. Examples are rivet nuts and clinch bolts, which serve as bolting points.

The strength properties and the chemical composition of rivet materials are laid down in numerous national and international standards. In the interests of avoiding electrochemical corrosion, it is advisable wherever possible to use materials of the same kind for rivets and component.

In mechanical engineering in general and in tank manufacture riveting has largely been displaced by welding.

Figure 6: Joint with punch rivet
a) Semi-hollow rivet,
b) Solid rivet.
1 Workpieces,
2 Punch rivet,
3 Riveting die,
4 Die plate.



Advantages and disadvantages compared with other joining techniques
- Unlike welding, riveting exerts no influence such as hardening or structural transformation on the material.
- No distortion of the components.
- Suitable for joining dissimilar materials.
- Riveting weakens the components.
- Butt joints cannot be riveted, and
- Riveting is generally more expensive than welding when performed outside the factory.

Automotive applications
- Riveting pivot/joint pins (power-window units, hinges, windshield-wiper linkages).
- Riveting reinforcement plates (in the course of repairs).

Punch riveting

Process

The punch-riveting technique joins solid materials employing stamping and riveting elements (solid or semi-hollow rivets) without piercing in a combined cutting and joining operation. The parts to be joined do not have to be pierced or predrilled, as is the case with other riveting techniques.

Punch riveting with semi-hollow rivets

The first stage in punch riveting with a semi-hollow rivet (Figure 6a) is to position the joining point of the components to be joined on the (bottom) die plate. The riveting die descends and presses the semi-hollow rivet through the upper sheet into the lower sheet in a single stamping operation. The rivet deforms and the bottom spreads to form a securing element, usually without fully penetrating the lower sheet.

Punch riveting with solid rivets

The first stage in punch riveting with a solid rivet (Figure 6b) is to position the joining point of the components to be joined on the (bottom) die plate. The top section of the rivet unit, including the blank holder, descends and the riveting die presses the rivet through the parts to be joined in a single stamping operation.

Equipment
Hydraulic joining equipment is used to produce the joints; in equipment of this type the riveting die and the die plate are arranged in a very rigid, C-shaped frame. The rivets can be supplied to the riveting tool loose or via magazined carrier strips.

Materials
The rivets must be harder than the materials to be joined. The most common materials are steel, stainless steel, copper and aluminum with various coatings.

Features
– Possible to join similar and dissimilar materials (e.g. steel, plastic or aluminum), parts of various thicknesses and strengths, and painted sheets.
– No preliminary piercing or drilling, no heat or vacuum extraction required.
– The total material thickness that can be riveted is 6.5 mm for steel and 11 mm for aluminum.
– The joining process generates minimal heat and noise.
– Tools have a long service life (approx. 300,000 rivet applications) and deliver consistent joint quality over a long period of time.
– High process reliability through monitoring of the process parameters.
– High forces have to be applied.
– Larger tong projections of the riveting tools are only possible under limited conditions owing to the rigidity requirements.

Applications
– Punch riveting with solid rivets: joining metal sheets in automotive engineering, e.g. power-window drives in passenger cars.
– Punch riveting with semi-hollow rivets: joining materials in passenger-car body manufacturing, white goods (household appliances), joining metals to composites (heat shields).

Penetration-clinching processes

Processes
Penetration clinching ("clinching") comprises mechanical joining processes which combine penetration, cold upsetting and sometimes also cutting in a single continuous joining operation without the application of additional heat. Based on this principle, it can be assigned to the process of joining by forming (see DIN 8593-5, [21]).

Distinctions are made between processes with and without cutting and those where the joining point is round or rectangular in shape.

"Tox clinching"
Some process variations are still referred to in technical parlance by the original manufacturer's designations, e.g. "Tox clinching" for penetration clinching with a round die without cutting (Figure 7b). The tools employed for "Tox clinching" are relatively small. The diameter can be varied to suit specific applications. A force-travel curve typical of "Tox clinching" can be broken down into five characteristic phases (A-E) (Figure 8).

Penetration clinching
Penetration clinching (Figure 7a) currently allows blanks up to 3 mm thick to be joined, whereby the total thickness of the two blanks should not exceed 5 mm. The blanks being joined can be of the same material (e.g. steel to steel) or dissimilar materials (e.g. steel to nonferrous metal). Penetration clinching can be used to join coated sheets and painted parts, as well as components to which adhesives have been applied. Multiple penetration clinching can be used to produce numerous elements to be joined (up to 50) in a single process (one stroke of the press, for example).

Advantages and disadvantages of penetration-clinching processes
- No need for an acoustic enclosure.
- Tox clinching to a large extent does not impair corrosion protection.
- When combined with cutting, there is a partial loss of corrosion protection.
- No distortion due to thermal stress.
- Painted, protected (oil, wax) and glued sheets can be clinched.
- Different materials can be joined (e.g. steel to plastic).
- Energy savings, since no power supply for welding and no cooling-water system are required.
- One side of the workpiece has a projection similar to that produced by a rivet head, while the other side has a corresponding depression.

Automotive applications
- Steel and aluminum bodies,
- Windshield-wiper brackets,
- Fastening interior door panels,
- Hinges, locks,
- Seat systems.

Figure 7: Penetration-clinching processes
a) Clinching,
b) "Tox clinching".
1 Die,
2 Parts to be joined,
3 Die plate.

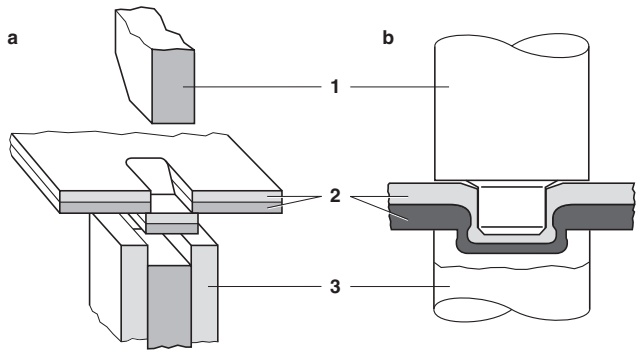
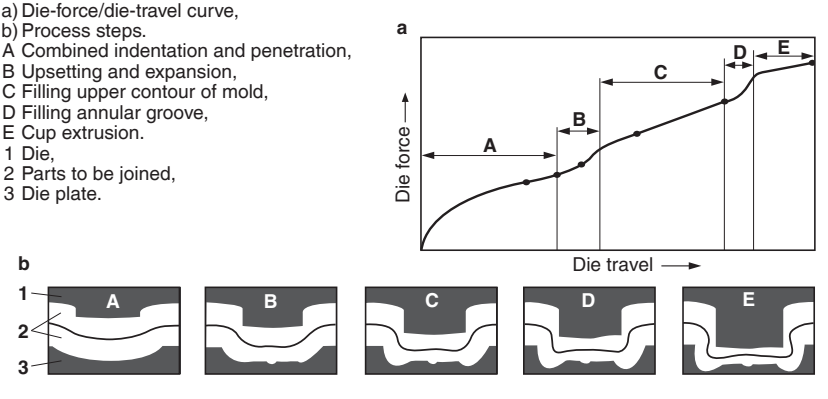


Figure 8: Typical force/travel curve in penetration-clinching processes
a) Die-force/die-travel curve,
b) Process steps.
A Combined indentation and penetration,
B Upsetting and expansion,
C Filling upper contour of mold,
D Filling annular groove,
E Cup extrusion.
1 Die,
2 Parts to be joined,
3 Die plate.



References

[1] Fügetechnik Schweißtechnik; 7th Edition, DVS Media, 2007.
[2] DIN-DVS Taschenbuch 284 – Schweißtechnik 7: Schweißtechnische Fertigung, Schweißverbindungen; 3rd Edition, DVS Media, 2009.
[3] DIN 1910-100:2008: Welding and allied processes – Vocabulary – Part 100: Metal welding processes with additions to DIN EN 14610:2005.
[4] DIN EN 14610:2005: Welding and allied processes – Definitions of metal welding processes; trilingual version EN 14610:2004.
[5] DIN-DVS Taschenbuch 283 – Schweißtechnik 6: Strahlschweißen, Bolzenschweißen, Reibschweißen; 4th Edition. DVS Media, 2009.
[6] DIN EN 1011: Welding – Recommendations for welding of metallic materials – Part 6 (2006): Laser beam welding; German version EN 1011-6:2005.
Part 7 (2004): Electron beam welding; German version EN 1011-7:2004.
[7] DIN-DVS Taschenbuch 196/1 – Schweißtechnik 5: Hartlöten; 5th Edition. DVS Media, 2008.
[8] DIN-DVS Taschenbuch 196/2 – Schweißtechnik 12: Weichlöten, gedruckte Schaltungen; 1st Edition. DVS Media, 2008.
[9] DIN ISO 857-2:2007: Welding and allied processes – Vocabulary – Part 2: Soldering and brazing processes and related terms (ISO 857-2:2005).
[10] DIN EN ISO 9454-1:2016: Soft soldering fluxes; classification and requirements; part 1: classification, labeling and packaging (ISO 9454-1:2016); German version EN ISO 9454-1:2016.
[11] DIN EN 1045:1997: Brazing – Fluxes for brazing – Classification and technical delivery conditions; German version EN 1045:1997.
[12] DIN EN ISO 9453:2014: Soft solder alloys – Chemical compositions and forms (ISO 9453:2014); German version EN ISO 9453:2014.
[13] DIN EN ISO 12224-1:1998: Solid wire, solid and flux cored – Specification and test methods – Part 1: Classification and requirements (ISO 12224-1:1997); German version EN ISO 12224-1:1998.
[14] DIN EN ISO 17672:2010: Brazing – Filler metals (ISO 17672:2010); German version EN ISO 17672:2010.
[15] DIN 660:2012: Round head rivets – Nominal diameters 1 mm to 8 mm.
[16] DIN 661:2011: Countersunk head rivets – Nominal diameters 1 mm to 8 mm.
[17] DIN 662:2011: Mushroom head rivets – Nominal diameters 1.6 mm to 6 mm.
[18] DIN 674:2011: Flat round head rivets – Nominal diameters 1.4 mm to 6 mm.
[19] DIN 7339:2011: Hollow rivets, one piece, draw from strip.
[20] DIN 7340:2011: Tubular rivets cut from the tube.
[21] DIN 8593-5:2003: Manufacturing processes joining – Part 5: Joining by forming processes; Classification, subdivision, terms and definitions.

Vehicle motion

Coordinate system

Vehicle motion is described by three translational and three rotational degrees of freedom (Figure 1). Generally, the system of coordinates (rectangular right-handed system) has its origin at the vehicle's center of gravity S. The x-axis (longitudinal axis) points forwards, i.e., in the direction of travel. It is situated in the plane that is perpendicular to the road surface. This plane is called the vehicle center plane. Situated perpendicular to the vehicle center plane is the y-axis (transverse axis), which points, when viewed in the direction of travel, to the left. The z-axis (vertical axis) points upwards (according to DIN ISO 8855 [1]).

Table 1: Quantities and units

Quantity		Unit
β	Float angle	degree
ψ	Yaw angle	degree
φ	Roll angle	degree
θ	Pitch angle	degree
a_x	Longitudinal acceleration	m/s²
a_y	Lateral acceleration	m/s²
a_z	Vertical acceleration	m/s²
a_t	Tangential acceleration	m/s²
a_c	Centripetal acceleration	m/s²
M_x	Rolling moment	N m
M_y	Pitching moment	N m
M_z	Yawing moment	N m
v_x	Longitudinal velocity	m/s
v_y	Lateral velocity	m/s
v_z	Vertical velocity	m/s

Degrees of freedom of vehicle motion

Translational motions

Definitions of motion variables

The translational motions are designated as follows:
– In x-direction: linear motion.
– In y-direction: lateral motion.
– In z-direction: lift.

Linear motion is caused by motive force, braking force, and frictional force. Lateral motion is a consequence of centrifugal force, crosswind force, and cornering force. Vertical motion is caused by wheel load and forces brought about by road-surface irregularities.

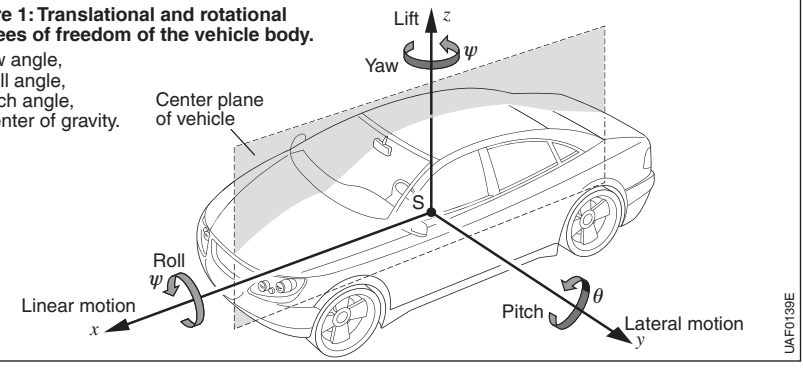
Translational velocities and accelerations

Single time derivation of the translational motion variables determines the linear, lateral, and vertical velocities (v_x, v_y, v_z) (quantities and units, see Table 1). Renewed derivation produces the linear acceleration a_x, lateral acceleration a_y, and vertical acceleration a_z.

Float angle

In the dynamics of lateral motion the center of gravity of the vehicle body does not always move along the x-axis. The angle that forms between the vehicle center plane and the instantaneous vehicle linear velocity (trajectory) is called the float angle β (Figure 2). It is counted from the

Figure 1: Translational and rotational degrees of freedom of the vehicle body.
ψ Yaw angle,
φ Roll angle,
θ Pitch angle,
S Center of gravity.



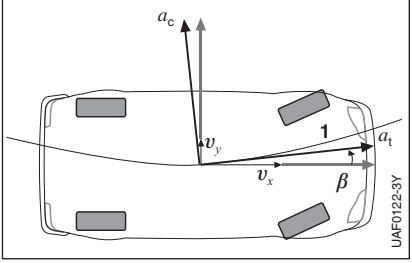
vehicle center plane to the trajectory. The float angle is calculated from the linear velocity v_x and the lateral velocity v_y in the following form:

$$\beta = \arctan \frac{v_y}{v_x}.$$

Because linear velocity during forward driving is by definition positive, the sign preceding the lateral velocity in this case also determines the sign preceding the float angle.

Centripetal acceleration and tangential acceleration
Acceleration is proportioned in the horizontal plane along the instantaneous vehicle trajectory into centripetal acceleration a_c and tangential acceleration a_t (Figure 2). The centripetal acceleration is that part which is perpendicular to the trajectory, while the tangential acceleration is that part in the tangential direction of the trajectory.

Figure 2: Float angle, centripetal and tangential acceleration
1 Trajectory.
β Float angle,
a_t Tangential acceleration,
a_c Centripetal acceleration,
v_x Longitudinal velocity,
v_y Lateral velocity.



Rotational motions
Definitions of motion variables
The following terms have been established to cover rotational motion.
– Yaw: Pure rotation about the z-axis with yaw angle ψ.
– Pitch: Pure rotation about the y-axis with pitch angle θ.
– Roll: Pure rotation about the x-axis with roll angle φ.

The signs preceding these rotations correspond to those of a rectangular right-handed system. The counting directions are shown in Figure 1. In many driving maneuvers the body simultaneously moves about more than one axis. Because, when viewed mathematically, the three rotations are not commutative, the following order for the rotations has been defined in DIN ISO 8855:
1. Yaw,
2. Pitch,
3. Roll.

Rotational velocities and accelerations
Single time derivation of the rotational motion variables produces the yaw, pitch and roll velocities. Renewed time derivation produces the yaw, pitch and roll accelerations.

Moments
Definitions
External forces acting on the vehicle can generate a moment about the center of gravity. This moment is broken down into three components.
– Portion of moment about the z-axis: Yaw moment M_z.
– Portion of moment about the y-axis: Pitching moment M_y.
– Portion of moment about the x-axis: Rolling moment M_x.

References
[1] DIN ISO 8855: Road vehicles – Vehicle dynamics and road-holding ability – Vocabulary.
(Earlier standard: DIN 70000).

Running resistances

Total running resistance

The running resistance F_W at constant driving speed v is the sum total of rolling resistance F_{Ro}, aerodynamic drag F_L, and climbing resistance F_{St} (Figure 1):

$$F_W = F_{Ro} + F_L + F_{St}.$$

In the case of an acceleration the acceleration resistance must additionally be overcome (see Dynamics of linear motion). This resistance is not taken into consideration in the following examination.

The power which must be transmitted through the drive wheels to overcome running resistance (running-resistance power) is:

$P_W = F_W\, v$ (quantity equation) or

$$P_W = \frac{F_W\, v}{3600}$$ (numerical-value equation).

The numerical-value equation is calculated with P_W in kW, F_W in N, v in km/h (quantities and units, see Table 1).

Individual running resistances

Rolling resistance

The rolling resistance F_{Ro} is the product of deformation processes which occur at the contact path between tire and road surface. A measure of the rolling resistance is the coefficient of rolling resistance f_R. The following equation applies to the rolling resistance:

$$F_{Ro} = f_R\, G \cos \alpha = f_R\, m\, g \cos \alpha .$$

The rolling resistance is determined by the coefficient of rolling resistance f_R and by the normal force, the force N acting vertically on the road surface ($G \cos \alpha$, see Figure 1).

On made-up roads the proportion of the rolling resistance caused by deformation is negligible. On roads with a soft or yielding surface, however, the deformation work makes up a significant proportion of the rolling resistance (Table 2).

The deformation work on the tire is created by flexion of the tire. These losses are influenced by the flexing amplitude, which increases with increasing wheel load and decreasing tire inflation pressure, and by the flexing frequency, which increases with the driving speed. The flexing frequency is also higher on tires with a small radius.

Table 1: Quantities and units

Quantity		Unit
E	Greatest cross-sectional area of vehicle (frontal area)	m^2
c_W	Drag coefficient	–
F	Motive force	N
f_K	Cornering resistance	N
F_L	Aerodynamic drag	N
F_{Ro}	Rolling resistance	N
F_{St}	Climbing resistance	N
F_W	Running resistance	N
f_K	Coefficient of cornering resistance	–
f_R	Coefficient of rolling resistance	–
G	Force due to weight, $G = m\, g$	N
G	Gravitational acceleration ≈ 9.81 m/s^2	m/s^2

Quantity		Unit
h	Height of projected distance l	m
l	Distance	m
m	Vehicle mass	kg
N	Normal force	N
P_L	Aerodynamic drag in kW	W
P_{St}	Hill-climbing power requirement	W
P_W	Running-resistance power	W
p	Gradient (tan α)	%
t	Time	s
v	Driving speed	m/s
v_0	Headwind speed	m/s
α	Incline angle, Gradient angle	°
ρ	Air density	kg/m^3

An approximate calculation of the rolling resistance can be made using the coefficients provided in Table 2 and Figure 2.

In turns, the rolling resistance is augmented by the cornering resistance.

$F_K = f_K \; G$.

The coefficient of cornering resistance f_K is a function of driving speed, curve radius, suspension geometry, tires, tire inflation pressure, and the vehicle's response under lateral acceleration.

Figure 1: Running resistance
α Gradient angle,
S Center of gravity,
F_L Aerodynamic drag, F_{Ro} Rolling resistance,
F_{St} Climbing resistance,
G Vehicle weight, N Normal force.

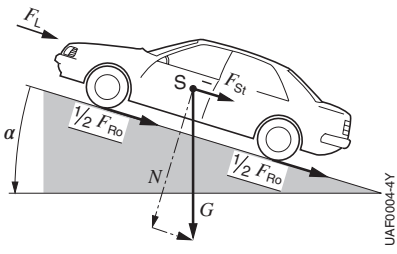
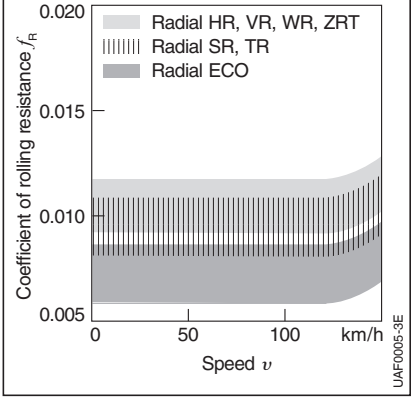


Figure 2: Coefficient of rolling resistance for radial car tires on smooth, level road surfaces under normal load and at correct tire inflation pressure



Aerodynamic drag

Aerodynamic drag is generated by air flowing round and through a vehicle as it is driving. It is calculated as a function of the air density ρ (ρ = 1.204 kg/m³ at sea level, normal pressure 1013 mbar, 20 °C), the cross-sectional area A of the vehicle, the drag coefficient c_w, and the square of the airflow velocity (sum total of driving speed v and headwind speed v_0) as:

$F_L = 0.5 \rho c_w A (v + v_0)^2$ or

$F_L = 0.0386 \rho c_w A (v + v_0)^2$

with v in km/h, F_L in N, ρ in kg/m³, A in m².

Typical value ranges for drag coefficient c_w and cross-sectional area A (frontal area) are listed for different passenger-car body configurations in Table 3. For commercial vehicles values for the drag coefficient of c_w = 0.4…0.9 and for the cross-sectional area of A = 4…9 m² are obtained.

The aerodynamic-drag power is

$P_L = F_L v = 0{,}5 \rho c_w A v (v + v_0)^2$ or

$P_L = 12.9 \cdot 10^{-6} c_w A v (v + v_0)^2$

with P_L in kW, F_L in N, v and v_0 in km/h, A in m², ρ = 1.204 kg/m³.

Table 2: Coefficients of rolling resistance

Road surface	Coefficient of rolling resistance f_R
Pneumatic car tires on	
large set pavement	0.011
small set pavement	0.011
concrete, asphalt	0.008
rolled gravel	0.02
unpaved road	0.05
field	0.1 - 0.35
Pneumatic truck tires on	
concrete, asphalt	0.006…0.01
Strake wheels in	
field	0.14…0.24
Track-type tractor in	
field	0.07…0.12
Wheel on rail	0.001…0.002

Table 3: Drag coefficients and aerodynamic drag for various vehicle body configurations

	Drag coefficient c_W	Frontal area in m^2	Aerodynamic drag in kW at a driving speed of	
			40 km/h	120 km/h
 Subcompact	0.29…0.37	2.05…2.20	0.5…0.7	13.2…18.1
 Compact class	0.22…0.32	2.18…2.28	0.4…0.6	10.7…16.2
 Medium class	0.23…0.35	2.20…2.38	0.4…0.7	11.3…18.5
 Station wagon	0.27…0.35	2.20…2.38	0.5…0.7	13.2…18.5
 Van	0.25…0.35	2.40…3.20	0.5…0.9	13.4…24.9
 Convertible				
- closed	0.28…0.38	1.94…2.20	0.4…0.7	12.1…18.6
- open	0.35…0.50	1.84…2.10	0.5…0.9	14.3…23.4
 Off-road vehicle	0.29…0.55	2.49…3.15	0.6…1.3	16.1…38.6
 Sports cars	0.27…0.40	1.65…2.20	0.4…0.7	10.0…19.6
 Luxury class	0.23…0.35	2.28…2.65	0.4…0.8	11.7…20.6

UAF0006-2E

Table information:
The mid-size class market segment features the lowest c_W values, followed by the luxury class, compact class, sports cars, closed convertibles, station wagons, vans, small cars up to the highest values for open-top convertibles and off-road vehicles.

However, the frontal area of a vehicle also has a significant influence on the power requirement for overcoming the drag. This changes the order with regard to the c_W value. Mid-size vehicles also require the lowest power here, followed by the compact class, sports cars and closed-top convertibles, which are nearly the same. This is followed by small vehicles and station wagons. Due to the size, luxury vehicles have fallen behind significantly and then open-top convertibles due to their high c_W value. Vans and off-road vehicles are found at the end.

This also shows that the aerodynamic drag at 40 km/h hardly differs between the classes. At 120 km/h, however, there are substantial differences because the power requirement increases with the speed cubed.

The largest cross-section of an automobile is:

$A \approx 0.9 \times$ lane width $\times$ height.

Table 3 shows values for the aerodynamic-crag power for different passenger-car body configurations at two different driving speeds. On account of the quadratic influence of the airflow velocity on the aerodynamic drag, the aerodynamic drag decisively determines the total running resistance at high speeds.

Empirical determination of coefficients for aerodynamic drag and rolling resistance

The vehicle is allowed to coast down in neutral under windless conditions on a level road surface. The time that elapses while the vehicle coasts down by a specific increment of speed is measured from two initial velocities, v_1 (high speed) and v_2 (low speed). This information is used to calculate the mean deceleration rates a_1 and a_2. The formulae and example from Table 4 are based on a vehicle weighing m = 1450 kg with a cross-section A = 2.2 m².

The method is suitable for application at driving speeds of less than 100 km/h.

Climbing resistance and downgrade force

Climbing resistance (F_{St} with positive sign) and downgrade force (F_{St} with negative sign) are calculated as:

$$F_{St} = G \sin\alpha = m\, g \sin\alpha$$

or, for a working approximation:

$$F_{St} \approx \frac{m\, g\, p}{100\ \%}.$$

This equation is valid for $p < 20$ % ($\alpha < 11$ °) gradient because at small angles:

$\sin\alpha \approx \tan\alpha$ (deviation below 2 % for $\alpha < 11$ °).

Refer to Table 5 for the climbing resistance, normalized to 1000 kg vehicle mass, as a function of the gradient p and the gradient angle α. "in" in the gradient scale in Table 5 means "divided by" and is customary in English-speaking countries.

Climbing power is calculated as:

$P_{St} = F_{St}\, v$ (quantity equation) or

$$P_{St} = \frac{F_{St}\, v}{3600} = \frac{m\, g\, v \sin\alpha}{3600}$$ (numerical-value equation)

with P_{St} in kW, F_{St} in N, and v in km/h. For a working approximation:

$$P_{St} = \frac{m\ g\ p\ v}{360{,}000}.$$

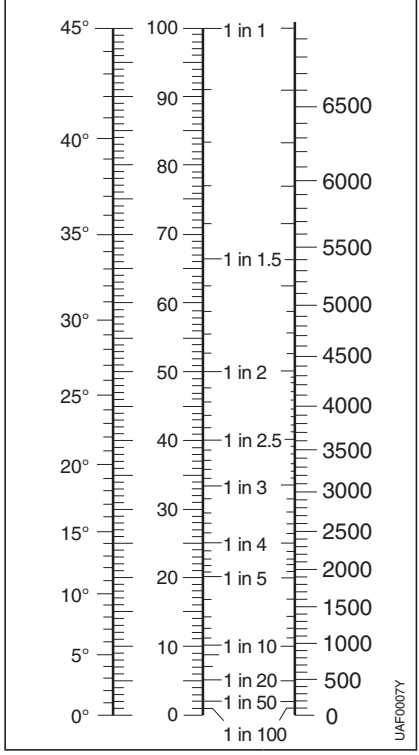
Table 4: Empirical determination of coefficients for aerodynamic drag and rolling resistance

	1st trial (high speed)	2nd trial (low speed)
Initial velocity Terminal velocity Interval between v_a and v_b	v_{a1} = 60 km/h v_{b1} = 55 km/h t_1 = 7.8 s	v_{a2} = 15 km/h v_{b2} = 10 km/h t_2 = 12.2 s
Mean velocity	$v_1 = \frac{v_{a1} + v_{b1}}{2}$ = 57.5 km/h	$v_2 = \frac{v_{a2} + v_{b2}}{2}$ = 12.5 km/h
Mean deceleration	$a_1 = \frac{v_{a1} - v_{b1}}{t_1}$ = 0.64 $\frac{\text{km/h}}{\text{s}}$	$a_2 = \frac{v_{a2} - v_{b2}}{t_2}$ = 0.41 $\frac{\text{km/h}}{\text{s}}$
Drag coefficient (all quantities as numerical values)	$c_W = \frac{6m(a_1 - a_2)}{A(v_1^2 - v_2^2)}$ = 0,29 (mit m = 1450 kg, A = 2.2 m²)	
Coefficient of rolling resistance (all quantities as numerical values)	$f_R = \frac{28.2(a_2 v_1^2 - a_1 v_2^2)}{10^3 \cdot (v_1^2 - v_2^2)}$ = 0.011	

Refer to Table 6 for the necessary climbing power, normalized to 1000 kg vehicle mass, as a function of the speed v and the climbing resistance F_{St} (values are limited to a vehicle power-to-weight ratio of 72 kW/t).

Table 5: Gradient angle and climbing resistance

Gradient angle α	Gradient p in %	Gradient	Climbing resistance F_{St} For m = 1000 kg in N



The gradient is:

$p = (h/l) \cdot 100\ \%$ or
$p = (\tan \alpha) \cdot 100\ \%$,

with h as the height of the projected distance l.

Example of calculating motive force and climbing power
To climb a hill with a gradient of p = 21%, a vehicle weighing 1,500 kg will require approximately 1.5 · 2 000 N = 3000 N of motive force at the wheels (value from Table 5) and at v = 40 km/h roughly 1.5 · 22 kW = 33 kW of climbing power (value from Table 6).

Table 6: Climbing resistance and climbing power

Values at m = 1000 kg					
Climbing resistance F_{St} in N	Climbing power P_{St} in kW at various speeds				
	20 km/h	30 km/h	40 km/h	50 km/h	60 km/h
6500	36	54	72	–	–
6000	33	50	67	–	–
5500	31	46	61	–	–
5000	28	42	56	69	–
4500	25	37	50	62	–
4000	22	33	44	56	67
3500	19	29	39	49	58
3000	17	25	33	42	50
2500	14	21	28	35	42
2000	11	17	22	28	33
1500	8.3	12	17	21	25
1000	5.6	8.3	11	14	17
500	2.3	4.2	5.6	6.9	8.3
0	0	0	0	0	0

Motor-vehicle dynamics

Characteristics of vehicle dynamics

The dynamics of vehicles can be divided into three subareas (Figure 1):
– dynamics of linear motion,
– dynamics of lateral motion, and
– dynamics of vertical motion.

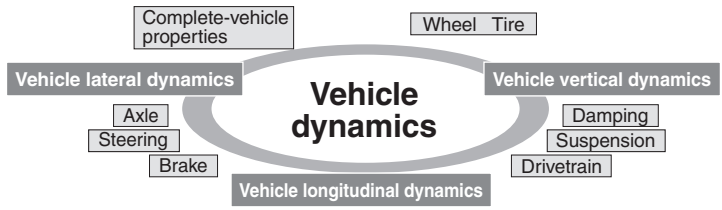
These three subareas can be roughly paraphrased as driving dynamics, agility, and low-frequency comfort without acoustics. Above all vehicle handling, which results from operation of the steering wheel, is considered in the dynamics of lateral motion. The dynamics of linear motion focuses on the acceleration behavior of the vehicle, i.e., all those situations in which the driver decelerates or accelerates. The dynamics of vertical motion concerns primarily the interaction with uneven roads and the effect of imbalance with rotating components, such as the wheel/tire system for example [1, 2, 3].

These three subareas should not be sharply distinguished from each other. There is for example the "Braking in a turn/curve" maneuver (see Driving-dynamics test procedures). This maneuver combines the dynamics of lateral motion with the dynamics of linear motion. Because this maneuver involves primarily the stability of the vehicle in the lateral direction and not deceleration values, it is assigned to the dynamics of lateral motion subarea. Another example the "tramp of a vehicle". This vehicle behavior combines the dynamics of lateral motion with the dynamics of vertical motion. Here an unwanted jerking motion is caused in the lateral direction which can arise when a wheel runs in a turn/curve over an uneven area and at the same time the damping is insufficient to maintain contact between tire and road.

Table 1: Quantities and units

Quantity		Unit
a	Acceleration, deceleration	m/s²
F	Motive force	N
F_L	Aerodynamic drag	N
F_{Ro}	Rolling resistance	N
F_{St}	Climbing resistance	N
F_W	Running resistance	N
f_R	Coefficient of rolling resistance	–
g	Gravitational acceleration ≈ 9.81 m/s²	m/s²
i	Overall transmission ratio between engine and drive wheels	–
k_m	Rational inertia coefficient	–
m	Engine torque	N m
m	Vehicle mass	kg
N	Engine speed	min^{-1}
k	Ratio of wheel loads	–
P	Engine power	W
P_A	Drive power	W
P_W	Motive power	W
r	Dynamic rolling radius of tire	m
s	Distance	m
t	Time	s
v	Driving speed	m/s
W	Work	J
α	Incline angle, Gradient angle	°
μ_r	Coefficient of static friction	–
η	Efficiency	–

Figure 1: Subareas of vehicle dynamics and the most important assemblies



Dynamics of linear motion

Features

The dynamics of linear motion (longitudinal dynamics) comprises the running resistances, the acceleration and deceleration, the forces and power outputs for drive and deceleration operations, and the energy expenditure required for these purposes, which is referred to as consumption (see Fuel consumption). The purpose of the motive and braking forces is to check the vehicle's dynamics of linear motion.

The change in a vehicle's motion caused by steering, acceleration and braking is brought about by forces which are transmitted to the road on account of friction between tire and road surface. The extent of the maximum transmissible force depends among other things on the force at the wheels acting on the road, the wheel load. The forces acting on the road are also influenced by speed-dependent aerodynamic lifting forces and downforces.

Motive force and engine speed

Vehicle with manual transmission

Motive force

The higher the engine torque M and overall transmission ratio i between the engine and drive wheels, and the lower the power-transmission losses, the higher the motive force F available at the driven wheels:

$$F = \frac{M\ i}{r} \cdot \eta \quad \text{oder} \quad F = \frac{P\ \eta}{v}.$$

η indicates the drivetrain efficiency. For a longitudinal engine $\eta \approx 0.88$ - 0.92, for a transverse engine $\eta \approx 0.91$ - 0.95.

The overall transmission ratio i for today's passenger cars is in the range of 1.5 to 20 and for off-road vehicles with additional reduction gears also much higher than this.

The motive force F is partially consumed in overcoming the running resistance F_W. Numerically higher transmission ratios are applied to deal with the substantially increased running resistance encountered on gradients (variable-speed transmission).

Vehicle and engine speeds

The following applies to the engine speed n in min^{-1} (rpm) with the dynamic rolling radius r of the tire in m and the driving speed v in m/s:

$$n = \frac{60\,v\,i}{2\,\pi\,r},$$

or with v in km/h:

$$n = \frac{1\,000\,v\,i}{2\,\pi\,60\,r}.$$

The driving speed v in m/s or in km/h can also be calculated from the engine speed according to the following formula:

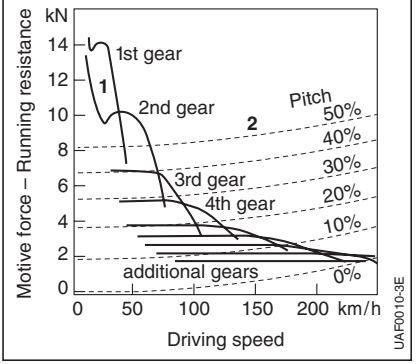
$$v\left[\frac{\text{m}}{\text{s}}\right] = \frac{n\,\pi\,r}{30\,i} \quad \text{oder} \quad v\left[\frac{\text{km}}{\text{h}}\right] = \frac{0{,}12\,n\,\pi\,r}{i}.$$

Vehicle with automatic transmission

When the formula for motive force is applied to automatic transmissions with hydrodynamic torque converters, the engine torque M is replaced by the torque at the converter turbine, while the rotational speed of the converter turbine is used in the formula for engine speed. The relationship between $M_{\text{Turb}} = f(n_{\text{Turb}})$ and the engine characteristic $M_{\text{Mot}} = f(n_{\text{Mot}})$ is determined using the characteristics of

Figure 2: Tractive-force running diagram for a car with automatic transmissions and hydrodynamic torque converter at full load

1 Motive-force lines (tractive-force lines) for different gears,
2 Running-resistance lines for different gradient (sum total of aerodynamic drag, rolling resistance and climbing resistance).



the hydrodynamic converter (see Hydrodynamic torque converter).

Tractive-force running diagram
The tractive force (motive force) at the wheels for the individual gears as a function of the driving speed can be read from the tractive-force running diagram (Figure 2). The kink typical of a hydrodynamic converter resulting from the end of torque multiplication can be seen in the first two gears. The converter is locked up by a clutch in higher gears. The top speed in each case as a function of gear and gradient can be determined from the points of intersection of the tractive-force lines with the running-resistance lines (sum total of rolling resistance, aerodynamic drag, and climbing resistance). To the left of the point of intersection the motive force is greater than the running resistance and the vehicle is consequently accelerated. At the point of intersection the motive force is utilized entirely to overcome the running resistance and the vehicle speed thus remains constant.

Acceleration
The surplus force $F - F_W$ accelerates the vehicle. Or decelerates it when F_W exceeds F:

$$a = \frac{F - F_W}{k_m m} \text{ oder } a = \frac{P\eta - P_W}{v k_m m}.$$

During a translational acceleration of the vehicle the rotating parts of the drivetrain (wheels, flywheel, crankshaft etc.) are also rotationally accelerated. The additional rotational resistance force required for this purpose is taken into account in the form of the rotational inertia coefficient k_m as the apparent increase in vehicle mass due to the rotating masses. The rotational inertia coefficient is dependent on the ratio of the vehicle mass m to the mass moment of inertia of the drivetrain reduced to the powered axle. The latter is quadratically influenced by the overall transmission ratio i between the engine and the drive wheels. The moment of inertia of the engine, which is approximately linear to the displacement V_H, also has a significant influence (Figure 3).

Adhesion to road surface
Coefficient of static friction
The coefficient of static friction μ_r (between the tires and the road surface) is determined by the driving speed, the condition of the tires, and the state of the road surface (Table 2). The figures cited apply to concrete and asphalt road surfaces in good condition. As a rule, they are in the region of $\mu < 1$. The coefficients of sliding friction (with wheel locked) are usually lower than the coefficients of static friction.

Tires for street-legal sports cars attain μ_r values of up to 1.3. Their rubber compounds are softer than in conventional tires and their treads are less pronounced. Tires which are used in motor sport and in which special soft rubber compounds are

Table 2: Coefficients of static friction for pneumatic tires on various road surfaces

Driving speed in km/h	Tire condition	State of road surface				
		Dry	Wet, water level approx. 0.2 mm	Heavy rainfall, water level approx. 1 mm	Puddles, water level approx. 2 mm	Iced (black ice)
		Coefficient of static friction μ_r				
50	new	0.85	0.65	0.55	0.5	0.1 and less
	worn[1]	1	0.5	0.4	0.25	
90	new	0.8	0.6	0.3	0.05	
	worn[1]	0.95	0.2	0.1	0.05	–
130	new	0.75	0.55	0.2	0	–
	worn[1]	0.9	0.2	0.1	0	

[1] Worn to 1.6 mm tread depth (legal minimum in Germany as per § 36 StVZO (German Road Traffic Licensing Regulations)).

used can provide friction coefficients of up to $\mu_r = 1.8$. These are for the most part smooth-tread slicks where the microgearing effects between rubber and road surface can be particularly pronounced by maximizing the tire contact patch.

Aquaplaning
Aquaplaning has a particularly dramatic influence on the contact between tire and road surface. It describes the state in which a layer of water separates the tire and the (wet) road surface (Figure 4). The phenomenon occurs when a wedge of water forces its way underneath the tire's contact patch and lifts it from the road. The tendency to aquaplane is dependent upon such factors as the depth of the water on the road surface, the vehicle's speed, the tread pattern, the tread wear, and the load pressing the tire against the road surface. Wide tires are particularly susceptible to aquaplaning. An aquaplaning vehicle cannot transmit control and braking forces to the road surface, and may as result go into a skid.

Accelerating and braking

The vehicle is regarded as accelerating or braking (decelerating) at a constant rate when acceleration a is constant. The equations set out in Table 3 apply to acceleration from a standstill and braking (deceleration) to a standstill ($v = 0$).

Maxima for acceleration and braking (deceleration)
When the motive or braking forces exerted at the vehicle's wheels reach such a magnitude that the tires are just still within their limit of adhesion (i.e., maximum adhesion is still present), the relationships shown in Tables 4 and 5 exist between the gradient angle α, coefficient of static friction μ_r, and maximum acceleration or deceleration. In this connection, k indicates the ratio between the load on driven or braked wheels and the total weight. If all wheels are driven or braked, $k = 1$. At 50 % load distribution $k = 0.5$. In practice, the figures that can be achieved are always somewhat lower, as all the vehicle's tires do not simultaneously exploit their maximum possible adhesion during each acceleration or deceleration. Electronic traction control and braking systems (TCS, ABS, ESC) maintain the traction level in the vicinity of the greatest possible coefficient of static friction μ_r.

It follows from the formula in Table 4 that on a level road surface with tires that exhibit $\mu_r < 1$ for all-wheel-driven vehicles (where $k = 1$) the maximum acceleration is $g(\mu_r - f_R)$ (f_R coefficient of rolling resistance).

Figure 3: Determining the rotational inertia coefficient k_m
V_H Engine displacement in liters,
m Vehicle mass,
i Overall ratio between engine and drive wheels,
r Wheel radius.
1 m/V_H = 500 kg/l,
2 m/V_H = 750 kg/l,
3 m/V_H = 1000 kg/l.

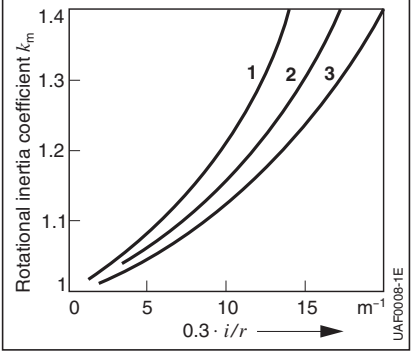
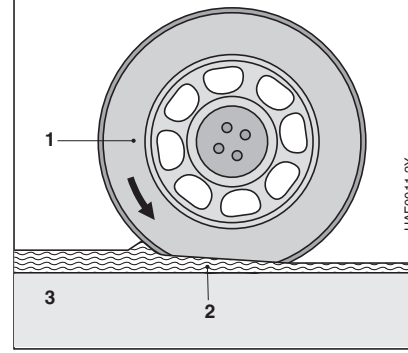


Figure 4: Aquaplaning
1 Tire, 2 Wedge of water, 3 Road surface.



Example:
$k = 0.5$; $g \approx 10$ m/s²,
$\mu_r = 0.6$; $p = 15$ %,
$a_{max} \approx 10 \cdot (0.5 \cdot 0.6 \pm 0.15)$ m/s²,
upgrade braking (+): $a_{max} = 4.5$ m/s²,
downgrade braking (–): $a_{max} = 1.5$ m/s².

Work and power
The power required to maintain a consistent rate of acceleration (deceleration) varies according to driving speed (Table 6). The power available for acceleration is:

Table 3: Accelerating and braking

	Equations for v in m/s	Equations for v in km/h
Acceleration or braking (deceleration) in m/s²	$a = \frac{v^2}{2s} = \frac{v}{t} = \frac{2s}{t^2}$	$a = \frac{v^2}{25.92s} = \frac{v}{3.6t} = \frac{2s}{t^2}$
Acceleration or braking time in s	$t = \frac{v}{a} = \frac{2s}{v} = \sqrt{\frac{2s}{a}}$	$t = \frac{v}{3.6a} = \frac{7.2s}{v} = \sqrt{\frac{2s}{a}}$
Accelerating or braking distance in m	$s = \frac{v^2}{2a} = \frac{vt}{2} = \frac{at^2}{2}$	$s = \frac{v^2}{25.92a} = \frac{vt}{7.2} = \frac{at^2}{2}$

Table 4: Acceleration and braking (deceleration)

	Level road surface	Inclined road surface α; $p = 100\,\% \cdot \tan\alpha$	
Limit acceleration or deceleration a_{max} in m/s²	$a_{max} = k\,g\,(\mu_r - f_R)$	$a_{max} = g\,(k(\mu_r - f_R)\cos\alpha \pm \sin\alpha)$ approximation[1]: $a_{max} \approx g\,(k(\mu_r - f_R) \pm 0.01p)$	+ when upgrade braking or downgrade acceleration – when upgrade acceleration or downgrade braking

[1] Valid to approx. $p = 20$ % (error under 2 %).

Table 5: Achievable acceleration a_e for given drive power P_A
a_e in m/s, P_A in kW, v in km/h, m in kg.

Level road surface	Inclined road surface	
$a_e = \frac{3600\,P_A}{k\,m\,v}$	$a_e = \frac{3600\,P_A}{k\,m\,v} \pm g\sin\alpha$	+ Downgrade acceleration – Upgrade acceleration for $g\sin\alpha$ applies with approximation[1] $g\,p$ with p in %

Table 6: Work and power

	Level road surface	Inclined road surface α; $p = 100\,\% \cdot \tan\alpha$ [%]	
Drive or braking work W in J, see[2]	$W = k\,m\,a\,s$	$W = m\,s\,(k\,a \pm g\sin\alpha)$ approximation[1]: $W = m\,s\,(k\,a \pm g\,p/100)$	+ Downgrade braking or upgrade acceleration – Downgrade acceleration or upgrade braking
Drive or braking power in W at speed v	$P_A = k\,m\,a\,s\,v$	$P_A = m\,v\,(k\,a \pm g\sin\alpha)$ approximation[1]: $P_A = m\,v\,(k\,a \pm g\,p/100)$	v in m/s. For v in km/h v/3.6 must be set.

[1] Valid to approx. $p = 20$ % (under 2 % error), [2] 1 J = 1 N m = 1 W s.

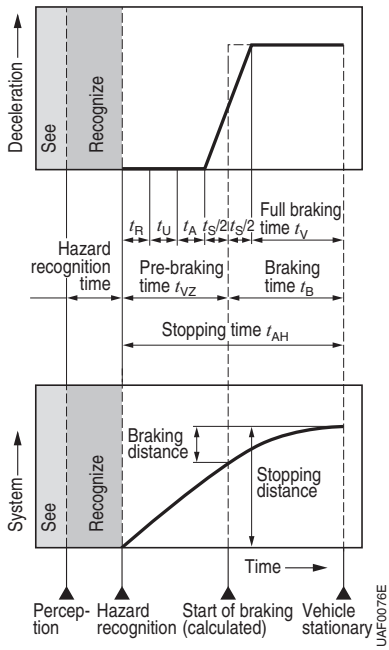
$P_a = P\,\eta - P_W$, where

P engine output,
η efficiency,
P_W motive power.

Reaction, braking and stopping
(in accordance with ÖNORM V 5050 [4] and [5])

Hazard recognition time
The hazard recognition time, also known as the danger reaction time, is the period of time that elapses between perceiving a visible obstacle or its movement and the time required to recognize it as a hazard (Figure 5; quantities, see Table 7). It is necessary for the driver to turn his eyes towards the hazardous situation, the hazard recognition and danger reaction time will extend by roughly 0.4 s. It is further increased by poor concentration or tiredness.

Figure 5: Actions: Reaction, braking and stopping
a) Deceleration while braking,
b) Distance covered while braking.



Prebraking time
The prebraking time (t_{VZ}) is the period of time that elapses between the moment the hazard is recognized and the start of braking, defined by way of calculation. Based on the following formula, the prebraking time is in the range from approx. 0.8 to 1.0 s:

$$t_{VZ} = t_R + t_U + t_A + \frac{t_S}{2}\,.$$

The reaction time (t_R) is the period of time that elapses between the moment a defined incitement to action occurs and the start of the first specifically targeted action. Instinctive hazard recognition triggers an inherent, automatic reaction (spontaneous reaction, involuntary reflex), enabling the vehicle driver to determine both the point of the reaction as well as the position of the reason for the reaction, delayed by the distance covered during the prebraking time. Human beings require about 0.2 s for the spontaneous reaction; however, the reaction time will be at least 0.3 s if the driver needs to make a decision to perform a preventive or evasive action in response to conscious hazard recognition (choice reaction, voluntary action).

The transfer time (t_U) is the period of time the driver requires to transfer the foot from the accelerator pedal to the brake pedal. The transfer time is in the range of about 0.2 s.

The response time (t_A) is the period of time it takes to transmit the pressure applied at the brake pedal via the brake system through to the point when the braking action becomes effective (complete build-up of the application force and incipient increase in vehicle deceleration).

Table 7: Quantities and units

Quantity		Unit
t_{VZ}	Prebraking time	s
t_R	Reaction time	s
t_U	Transfer time	s
t_A	Response time	s
t_S	Pressure build-up time	s
t_B	Braking time	s
t_V	Full braking time	s
t_{AH}	Stopping time	s
s_{AH}	Stopping distance	m

The pressure build-up time (t_S) is the period of time that elapses between the braking action taking effect and reaching fully effective braking deceleration. Alternatively, half the pressure build-up time ($t_S/2$) may be assumed as the start of braking, determined by way of calculation.

In accordance with the EU Council of Ministers Directive EEC 71/320 Addendum 3/2.4 [6], the sum of the response time and pressure build-up time may not exceed 0.6 s. A poorly maintained braking system will lengthen the response and pressure build-up time.

Braking time
The braking time (t_B) is the period of time that elapses between the mathematically calculated start of braking and the moment the vehicle comes to a complete stop. This period of time comprises half the pressure build-up time ($t_S/2$) (calculation assumption: only half the pressure build-up time but complete braking deceleration), as well as the full braking time (t_V), during which maximum braking deceleration is actually effective.

$$t_B = \frac{t_S}{2} + t_V .$$

Stopping time and stopping distance
The stopping time (t_{AH}) is the sum of the prebraking time (t_{VZ}) and braking time (t_B).

$$t_{AH} = t_{VZ} + t_B .$$

The stopping distance (s_{AH}) can be calculated by way of integration (Tables 8 and 9).

Safety margin
The minimum safety margin must be equal to the distance covered during the prebraking time t_{VZ}. The figure for a prebraking time of t_{VZ} = 1.08 s for speeds in km/h would be $0.3v$ meters; however, a minimum of $0{,}5v$ meters is advisable outside of built-up areas.

Passing (overtaking)
The complete passing maneuver involves pulling out of the lane, overtaking the other vehicle, and returning to the original lane (Figure 6). Passing can take place under a wide variety of very different circumstances and conditions, so precise calculations are difficult. For this reason, the following calculations, graphs, and illustrations will confine themselves to an

Table 8: Stopping time and stopping distance
a in m/s², v in m/s, t_{vz} in s.

	Equations for v in m/s	Equations for v in km/h
Stopping time t_{AH} in s	$t_{AH} = t_{VZ} + \frac{v}{a}$	$t_{AH} = t_{VZ} + \frac{v}{3.6\,a}$
Stopping distance s_{AH} in m	$s_{AH} = v\,t_{VZ} + \frac{v^2}{2\,a}$	$s_{AH} = \frac{v}{3.6}\,t_{VZ} + \frac{v^2}{25.92\,a}$

Table 9: Stopping distance as a function of driving speed and deceleration

Deceleration a in m/s²	Driving speed prior to braking in km/h												
	10	30	50	60	70	80	90	100	120	140	160	180	200
	Distance during prebraking time (delay) of 1 s in m												
	2.8	8.3	14	17	19	22	25	28	33	39	44	50	56
	Stopping distance in m												
4.4	3.7	16	36	48	62	78	96	115	160	210	270	335	405
5	3.5	15	33	44	57	71	87	105	145	190	240	300	365
5.8	3.4	14	30	40	52	65	79	94	130	170	215	265	320
7	3.3	13	28	36	46	57	70	83	110	145	185	230	275
8	3.3	13	26	34	43	53	64	76	105	135	170	205	250
9	3.2	12	25	32	40	50	60	71	95	125	155	190	225

examination of two extreme conditions: passing at a constant velocity and passing at a constant rate of acceleration. We can simplify graphic representation by treating the passing distance s_u as the sum of two (straight-ahead) components, while disregarding the extra travel involved by pulling out of the lane and back in again.

Passing distance
This passing distance is

$s_U = s_H + s_L$ (variables in Table 10).

The distance s_H which the more rapid vehicle must cover compared to the slower vehicle (considered as being stationary) is the sum of the vehicle lengths l_1 and l_2 and the safety margins s_1 and s_2.

$$s_H = s_1 + s_2 + l_1 + l_2 .$$

During the passing time t_u the slower vehicle covers the distance s_L which the faster vehicle has to travel in addition to s_H so that the safety margin is maintained (s_L in m, t_u in s):

$$s_L = \frac{t_U v_L}{3{,}6} \quad \text{(for } v \text{ in km/h).}$$

Passing at constant speed
On highways with more than two lanes, the overtaking vehicle will frequently be traveling at a speed adequate for passing before the actual process begins. The passing time (from initial lane change until return to the original lane has been completed) is then:

$$t_U = \frac{3{,}6\, s_H}{v_H - v_L}$$

with t in s, s in m and v in km/h, and the passing distance is

$$s_U = \frac{t_U v_H}{3{,}6} \approx \frac{s_H v_H}{v_H - v_L} .$$

Passing with constant acceleration
On narrow roads, the vehicle will usually have to slow down to the speed of the preceding car or truck before accelerating to pass. The attainable acceleration figures depend upon engine output, vehicle mass, speed, and running resistance. These generally lie within the range of 0.4 to 0.8 m/s², with up to 1.4 m/s² available in lower gears for further reductions in passing time. The distance required to complete the passing maneuver should never exceed half the visible stretch of road.

Operating on the assumption that a constant rate of acceleration a can be maintained for the duration of the passing maneuver, the passing time will be

$$t_U = \sqrt{\frac{2 s_H}{a}} .$$

The distance which the slower vehicle covers within this period is defined as $s_L = t_U v_L/3.6$; hence, the passing distance is:

$$s_U = s_H + \frac{t_U v_L}{3{,}6} ,$$

with t in s, s in m and v in km/h.

The left side of Figure 7 shows the relative distances s_H for speed differentials $v_H - v_L$ and acceleration rates a, while the right side shows the distances s_L covered by the vehicle being passed at various

Figure 6: Passing distance
Quantities, see Table 14.
(For clarity reasons s_L is depicted small here, in practical cases $s_L > s_H$.)
1 Passed vehicle,
2 Passing vehicle.

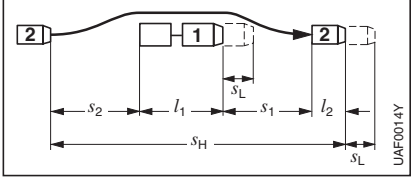


Table 10: Quantities and units

Quantity		Unit
a	Acceleration	m/s²
l_1, l_2	Vehicle length	m
s_1, s_2	Safety margin	m
s_H	Relative distance traveled by passing vehicle	m
s_L	Distance traveled by vehicle	m
s_U	Being passed	m
t_U	Passing distance	s
v_L	Passing time Speed of slower vehicle	km/h
v_H	Speed of faster vehicle	km/h

speeds v_L. The passing distance s_U is the sum of s_H and s_L.

First, determine the distance s_H to be traveled by the passing vehicle. Enter this distance on the left side of the graph between the Y axis and the applicable ($v_H - v_L$) line or acceleration line. Then extrapolate the line to the right, over to the speed line v_L.

Example 1 (constant acceleration):
$v_L = v_H = 50$ km/h (v at start of overtaking),
$a = 0.4$ m/s²,
$l_1 = 10$ m, $l_2 = 5$ m,
$s_1 = s_2 = 0.3\ |v_L| = 0.3\ |v_H| = 15$ m,
$\rightarrow s_H = 45$ m.

Solution
Enter intersection of $a = 0.4$ m/s² and
$s_H = 45$ m
in the left side of the graph.

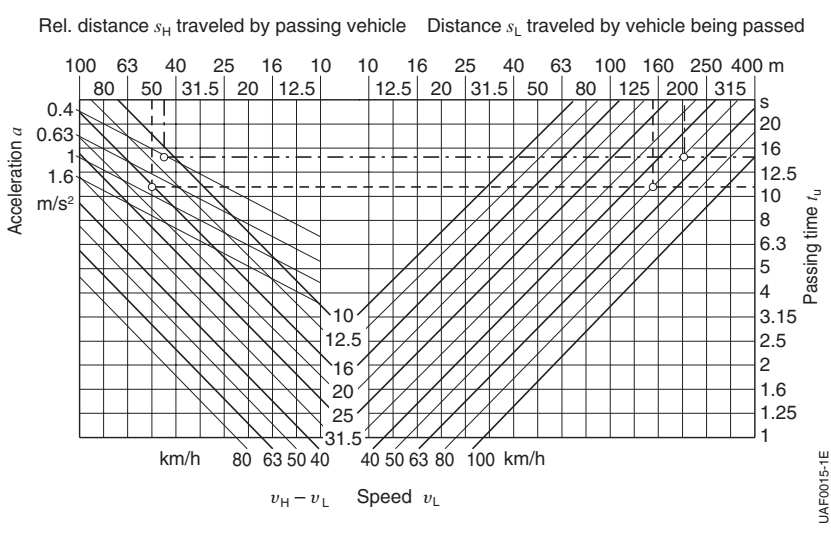
Indication: $t_U = 15$ s, $s_L = 210$ m.
The following applies: $s_U = s_H + s_L = 255$ m.

Example 2 (constant speed):
$v_L - v_H = 16$ km/h,
$v_H = 66$ km/h,
$v_L = 50$ km/h,
$s_1 = 0.3\ |v_L| = 15$ m,
$s_2 = 0.3\ |v_H| = 20$ m,
$\rightarrow s_H = 50$ m.

Solution
Intersection of $v_L - v_H = 16$ km/h with
$s_H = 50$ m
in the left side of the graph.
Indication: $t_U = 11$ s, $s_L = 150$ m.
The following applies: $s_U = s_H + s_L = 200$ m.

Visual range
For safe passing on narrow roads, the visibility must be at least the sum of the passing distance plus the distance which would be traveled by an oncoming vehicle while the passing maneuver is in progress. This distance is approximately 400 m if the vehicles approaching each other are traveling at speeds of 90 km/h, and the vehicle being overtaken at 60 km/h.

Figure 7: Graph for determining passing distance



Dynamics of lateral motion

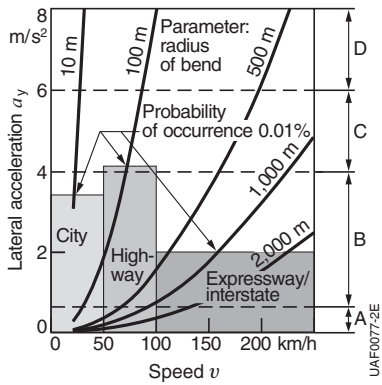
Features

The dynamics of lateral motion deal with all the vehicle states in which a dynamic response occurs in the transverse direction to the vehicle longitudinal axis. These maneuvers are caused above all by the steering of the vehicle, but also by external influences such as crosswind.

When it comes to the dynamics of lateral motion, the primary focus is on firstly a vehicle response deemed to be appropriate by the driver on account of a steering input and secondly the stability of the vehicle. This behavior is determined by three properties of the vehicle:

- Properties of the complete vehicle: total mass, center of gravity, aerodynamics.
- Properties of the axles: kinematics and elastokinematics of the axle, behavior of the tires and brakes, and spring and shock-absorber setting.
- Properties of the controllers: electric steering assistance and driving-dynamics control.

Figure 8: Lateral-acceleration ranges
A Small-signal range,
B Linear range (relevant to normal car drivers),
C Transition range,
D Limit range (emphasis on stability, relevant to press and experts).



Ranges of lateral acceleration

Passenger cars can reach lateral accelerations of up to 10 m/s^2. This range is below sports racing cars, which can reach up to 40 to 50 m/s^2, but significantly above commercial vehicles, which depending on size and load can only achieve 4 to 5 m/s^2. Lateral acceleration is subdivided into the following ranges (Figure 8):

Small-signal range

The range from 0 to 0.5 m/s^2 is known as the small-signal range. The phenomenon to be considered in this range is the straight-running behavior, caused by crosswind and irregularities in the road, such as ruts. Crosswind excitation is induced by blustery, gusty wind and by driving into and out of wind shadow areas.

Table 11: Quantities and units

Quantity		Unit
δ	Axle steering angle (steering angle)	rad
δ_H	Steering-wheel angle	rad
α_v	Slip angle of front axle	rad
α_h	Slip angle of rear axle	rad
β	Float angle	rad
ψ	Yaw angle	rad
ω_e	Undamped natural frequency	s^{-1}
θ	Yaw moment of inertia	Nms^2
ρ	Air density	kg/m^3
τ	Angle of impact	rad
a_y	Lateral acceleration	m/s^2
c_S	Lateral-force coefficient	–
c_M	Yaw-moment coefficient	–
i_l	Steering ratio	–
l	Wheelbase	m
l_v	Distance between front axle and center of gravity	m
l_h	Distance between rear axle and center of gravity	m
m	Total mass (weight)	kg
v	Longitudinal velocity	m/s
v_r	Resulting wind impact velocity	m/s
v_W	Wind velocity	m/s
A	Frontal area	m^2
C_v	Rear cornering stiffness Front axle	N/rad
C_h	Rear cornering stiffness Rear axle	N/rad
D	Damping factor	1/rad
F_{SV}	Lateral force on the front axle	N
F_{SH}	Lateral force on the rear axle	N
F_{FL}	Centrifugal force	N
F_S	Crosswind force	N
M_Z	Crosswind yaw moment	Nm

Linear range
The range from 0.5 to 4 m/s^2 is known as the linear range, as the vehicle behavior that occurs in this range can be described with the aid of the linear single-track model. Typical maneuvers involving dynamics of lateral motion include sudden steering input, changing driving lanes as well as combinations of maneuvers involving dynamics of both lateral and longitudinal motion, such as load-change reactions when cornering.

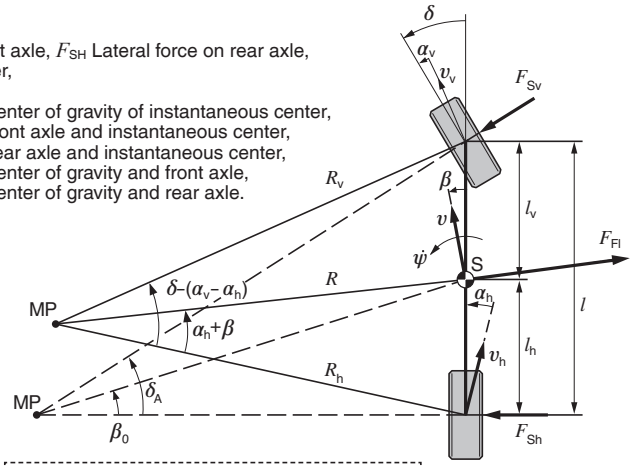
Transition range.
In the lateral acceleration range from 4 to 6 m/s^2, depending on their design features, passenger vehicles are categorized as either still linear or already nonlinear. This range is therefore considered to be a transition range. In this range, vehicles with maximum lateral acceleration of 6 to 7 m/s^2 (e.g., true offroad vehicles) already feature nonlinear characteristics, while vehicles that achieve higher levels of lateral acceleration (e.g., sports cars) still behave in line with linear characteristics.

Limit range
The lateral acceleration range above 6 m/s^2 is reached only in extreme situations and is therefore referred to as the limit range. In this range, the vehicle characteristics are predominantly nonlinear, with the main emphasis shifted to vehicle stability. This range is reached on racing circuits or in situations leading up to accidents in normal road traffic.

The average driver generally drives in the range up to 4 m/s^2. This means both the small signal range as well as the linear range are relevant to the car driver when subjectively assessing the situation (Figure 8). For the average car driver, the probability of lateral acceleration occurring decreases exponentially with lateral acceleration.

Figure 9: Single-track model for steady-state circular course driving

β Float angle,
$\dot{\psi}$ Yaw rate,
β_0 Float angle with wheel rolling free of slip,
δ_A Ackermann angle,
δ Steering angle,
α_v Slip angle, front wheel, α_h Slip angle, rear wheel,
v_v Tire velocity at front axle,
v_h Tire velocity at rear axle,
l Wheelbase,
F_{Fl} Centrifugal force,
F_{Sv} Lateral force on front axle, F_{SH} Lateral force on rear axle,
MP Instantaneous center,
S Center of gravity,
R Distance between center of gravity of instantaneous center,
R_v Distance between front axle and instantaneous center,
R_h Distance between rear axle and instantaneous center,
l_v Distance between center of gravity and front axle,
l_h Distance between center of gravity and rear axle.



Fast skidpad:
The wheels rolls with lateral slip → this results in slip angles and therefore also in lateral forces.

Slow skidpad:
The wheels roll without lateral slip → no slip angles and therefore also no lateral forces.

UAF0078-4Y

Linear single-track model

Features

Important deductions relating to the dynamic characteristics of lateral motion can be gained from the linear single-track model. The single-track model combines the lateral dynamic properties of one axle and its wheels to form one effective wheel [3]. In the simplest version, as illustrated here, the characteristics under consideration are positioned in the linear range. This means, for example, the lateral tire force is proportional to the slip angle, or the lateral aerodynamic force is proportional to the angle of impact. This model version is therefore called the linear single-track model. The nonlinear behavior of the complete vehicle in the lateral-acceleration range of 0.5 to 4 m/s^2 can already be described very well with these linear properties. The most important model assumptions are:

– Kinematics and elastokinematics of the axle are only modeled linearly,
– The lateral force buildup of the tire is linear and the aligning torque is ignored,
– The center of gravity is assumed to be at road level. The only rotational degree of freedom of the vehicle is therefore the yaw motion. Rolling, pitching and bounce are not taken into consideration.

With the assumptions of the linear single-track model the vehicle handling is reduced to two translational and one rotational degree of freedom of the vehicle body.

Self-steering effect

Figure 9 represents the single-track model in connection with the fast and slow skidpad. The planar motion of this rigid body can be interpreted as a rotation about the instantaneous center MP. The results in the following for the speed at the center of gravity S:

$$v = \dot{\psi} R .$$

This representation results in the following interrelationships describing the kinematics of slip angles [1, 2, 3]:

$$\alpha_v = \delta - \beta - \frac{\dot{\psi}\, l_v}{v} ; \quad \alpha_h = -\beta - \frac{\dot{\psi}\, l_h}{v} .$$

Together with the torque balance, it is possible to calculate the change in the steering-wheel angle in connection with increasing lateral acceleration a_y for the skidpad maneuver at a constant radius. This results in the definition of the self-steering gradient (quantities, see Table 11):

$$EG = \frac{d\delta}{da_y} = \frac{m}{l} \left(\frac{l_h}{C_v} - \frac{l_v}{C_h} \right) .$$

All passenger vehicles are designed so that they understeer in the linear lateral acceleration range. The EG value for passenger vehicles is in the range of about 0.25 degrees $\cdot$ s^2/m.

In terms of the dynamics of lateral motion, the self-steering gradient characterizes the stability and damping of the vehicle. In addition, the significance of the self-steering gradient for the average car driver becomes apparent in that the steering angle requirement increases the faster the cornering speed. This draws the driver's attention to the increasing lateral acceleration.

The float-angle gradient (SG) can be calculated from Figure 9. The float angle gradient should be as low as possible in order to increase the stability of the vehicle [1, 2, 3].

$$SG = \frac{d\beta}{da_y} = \frac{m}{C_h} \frac{l_v}{l} .$$

The SG value range for passenger cars is below 0.35 degrees $\cdot$ s^2/m. Greater float-angle gradients are perfectly manageable for experts, but already for stationary maneuvers give rise to a bad safety feeling for regular drivers and are therefore avoided.

Yaw gain

The yaw gain defines the degree of yaw response that a vehicle executes in response to a steering angle in the quasi-steady-state range. The yaw gain factor can be determined by conducting the follow test procedure: When driving at a constant speed, the steering wheel is turned with sinusoidal motion at a frequency of less than 0.2 Hz. The steering-angle amplitude is selected in order to achieve a maximum lateral acceleration of about 3 m/s^2.

Starting at a speed of 20 km/h, the maneuver is repeated at a speed that is increased by 10 km/h each time. Provided no aerodynamic influences occur at high speeds (lift or upthrust forces at the front and rear axles), the test will produce yaw gain curves that essentially agree with the following equation, derived from the linear single-track model [1, 2, 3]:

$$\left(\frac{\dot{\psi}}{\delta}\right)_{\text{stat}} = \frac{v}{l + EG\ v^2}$$

Figure 10 shows the yaw gain for a vehicle that tends to oversteer (EG < 0), that has neutral steering (EG = 0), and that understeers (EG > 0). Only the vehicle that understeers is acceptable at high speeds and therefore has the right vehicle dynamics even when driving in a straight line. The speed at which a vehicle that tends to understeer has the greatest yaw response is known as the characteristic speed v_{char}. In the linear single-track model, this speed is expressed as:

$$v_{\text{char}} = \sqrt{\frac{l}{EG}}$$

Damping factor

The following equilibrium of forces in the lateral direction is derived for the linear single-track model:

$$m\ a_{\text{y}} = F_{\text{SV}} \cos\delta + F_{\text{SH}}\ .$$

For the torque balance:

$$\theta\ \ddot{\psi} = F_{\text{SV}}\ l_{\text{v}} \cos\delta + F_{\text{SH}}\ l_{\text{h}}\ .$$

The damping factor D for excitation in terms of the dynamics of linear motion can be derived from the two equations:

$$D = \frac{1}{\omega_{\text{e}}} \left(\frac{C_{\text{v}} + C_{\text{h}}}{m\ v} + \frac{C_{\text{v}}\ l_{\text{v}}^2 + C_{\text{h}}\ l_{\text{h}}^2}{\theta\ v} \right)\ .$$

The following equation expresses the undamped natural frequency:

$$\omega_{\text{e}} = \sqrt{\left(\frac{C_{\text{h}}\ l_{\text{h}} - C_{\text{v}}\ l_{\text{v}}}{\theta} + \frac{C_{\text{v}}\ C_{\text{h}}\ l^2}{\theta\ m\ v^2} \right)}\ .$$

The damping factor of a vehicle can be identified, for example, from the yaw response of sudden steering or step input. The vehicle is designed to ensure that damping is as high as possible.

Figure 11 shows the damping factor and yaw gain for various self-steering gradients. This results in the following conflict of objectives:

– A high self-steering gradient is required if the vehicle is to have good straight-running characteristics,
– The self-steering gradient must be as low as possible to facilitate a high damping factor, particularly at high speeds.

Figure 10: Speed-dependent yaw gain

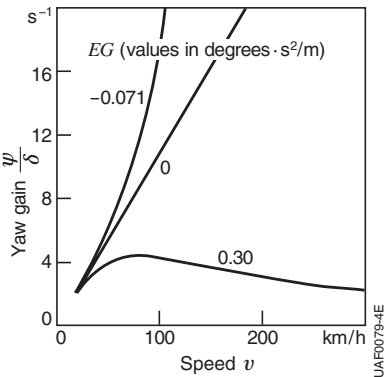
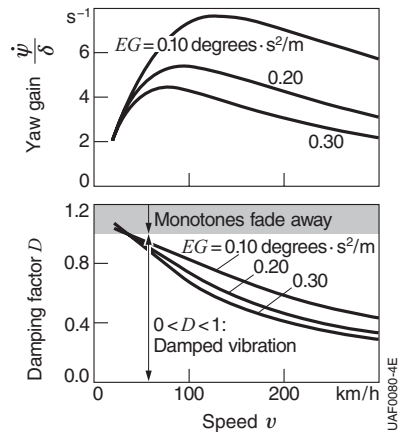


Figure 11: Damping factor and yaw gain



Lateral-agility diagram
A further important variable governing vehicle balance is the total steering ratio i_l. The steering-wheel angle, together with the total steering ratio i_l, is calculated from the axle steering angle:

$$\delta_H = i_l \ \delta .$$

This results in the following equation for maximum yaw gain [1]:

$$\left(\frac{\dot{\psi}}{\delta_H}\right)_{max} = \frac{1}{2\, i_l \sqrt{l\ EG}} .$$

This maximum is plotted in the lateral-agility diagram (Figure 12) as a function of the steering ratio. The diagram additionally shows the EG isolines. The self-steering gradient is constant along these curves. The desired target ranges for the yaw gain and steering ratio can be plotted in this diagram for the purpose of defining the necessary self-steering gradients.

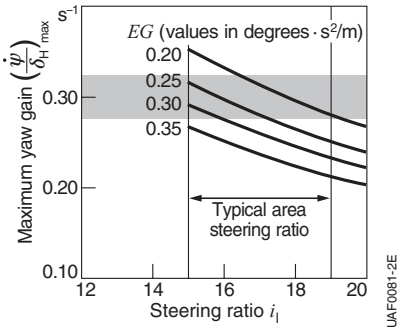
If only the steering ratio is changed in a vehicle, the maximum yaw gain can be determined in the lateral-agility diagram by shifting the baseline along the EG isolines. The shift will be along the vertical axis if the axis characteristics are varied.

General single-track and two-track model
The linear single-track model can be used to specify analytically the self-steering gradient, the float-angle gradient, and the yaw gain with associated damping factor; this makes the relationships clear. If the desire is to describe real vehicle behavior more accurately, more and more nonlinearities must be taken into account. This can be realized for example by taking into account the nonlinear behavior of the lateral force from the slip angle. With such nonlinearities it is possible to develop single-track models which also describe the roll of the vehicles and the handling up to a steering frequency of 3 Hz and are also valid in the driving-dynamic limit range [7].

Almost exclusively so-called two-track models are used in the automotive industry to predict the dynamics of lateral motion. The big advantage of these two-track models lies in the component-oriented modeling of the complete vehicle and thus better integration in the development sequence.

The use of two-track models is also sensible for the following aspects “Dynamics of lateral motion caused by crosswind” and “Roll behavior when cornering”.

Figure 12: Lateral-agility diagram
EG Self-steering gradient.



Dynamics of lateral motion caused by crosswind
Wind can induce lateral dynamics in motor vehicles. The vehicle responds to this external influence by drifting off course, lateral acceleration, and a change in yaw angle and roll angle. The driver then attempts to take corrective action to counteract this discrepancy. Consequently, the response capabilities of the driver as well as the correctability of the vehicle are taken into consideration in a second stage. According to current findings, the direct response of the vehicle to crosswind is the principal variable for subjectively assessing overall vehicle stability in crosswinds. This offers the advantage that the interaction between crosswind and vehicle response can be effectively observed by analysis.

Characteristically, the average car driver perceives two states induced by wind excitation:
– Natural crosswind, which can vary in terms of direction and wind velocity while driving,
– Driving into and out of wind shadow areas where forces of greatly varying strength can act on the vehicle.

The automotive industry strives to minimize the effects of excitations triggered by wind forces by taking the following vehicle factors into consideration:
– "Cornering stiffness" of the tires, i.e., to what extent the lateral force changes as the slip angle increases. The wheel load of the tire remains constant in this consideration,
– Total mass (weight) of the vehicle,
– Center of gravity,
– Axle characteristics,
– Equilateral and reciprocal suspension,
– Damping,
– Kinematics and elastokinematics of the axles,
– Aerodynamic shape and frontal area of the vehicle.

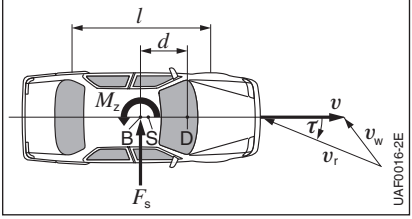
Aerodynamic forces and moments

When a vehicle moves at velocity v in a wind at a speed v_w, the vehicle will be subject to wind impact applied at the resulting velocity v_r. In connection with natural crosswind, the angle of impact τ generally differs from 0 degrees and therefore generates a lateral force F_S and a yawing moment M_Z that acts on the vehicle.

In aerodynamics [8] it is standard practice to specify dimensionless coefficients instead of forces and moments. Therefore:

$$F_S = c_S \frac{\rho}{2} v_r^2 A \; ; \; M_Z = c_M \frac{\rho}{2} v_r^2 A l .$$

Figure 13: Vehicle exposed to crosswind
D Center of pressure point, S Center of gravity, B Aerodynamic reference point.
v Driving speed,
v_w Wind velocity,
v_r Resulting velocity,
τ Angle of impact, F_S Crosswind force,
l Wheelbase, d Distance between B and D,
M_z Yawing moment.



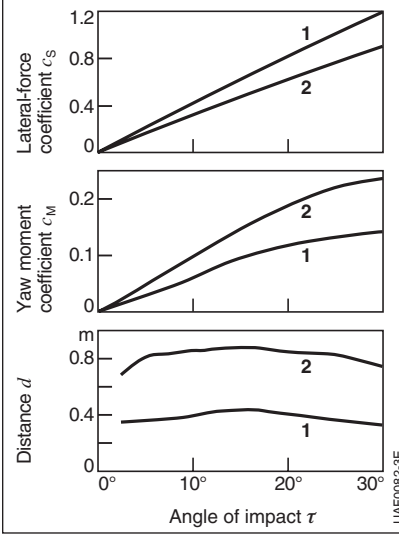
The moment M_Z and the lateral force F_S, which is defined at the mid-point of the wheelbase, can be represented by a single lateral force F_S when the point of impact is located at the center of the pressure point D (Figure 13). The distance d between the aerodynamic reference point B and the center of the pressure point D is calculated as follows:

$$d = \frac{M_z}{F_s} = \frac{c_M \; l}{c_S} \; .$$

To keep aerodynamic influences as low as possible, appropriate steps should be taken to ensure that the center of the pressure point D is as close to the vehicle's center of gravity S as possible. This consequently reduces the effective impact of the moment.

Figure 14 represents the aerodynamic coefficients of the two most typical structural shapes of vehicles, i.e., the station wagon and sedan, as a function of the angle of impact τ. The resulting distance d is considerably lower for stations wagons than for sedans (Figure 13). For vehicles

Figure 14: Lateral-force coefficient and pressure point
1 Station wagon, 2 Sedan.
d Distance between aerodynamic reference point B and center of pressure point D.

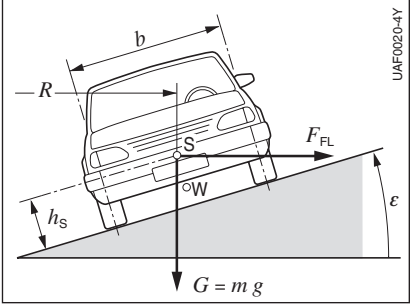


with the center of gravity located at the mid-point of the wheelbase, the station wagon structure is therefore less sensitive to crosswind than the sedan structure.

Roll behavior when cornering

When corning the vehicle body rolls about its roll axis. This rolling motion is caused by the centrifugal force F_{FL} (see Figures 9 and 15), which is applied at the vehicle's center of gravity. The magnitude of this rolling motion depends upon the rates of the springs and their response to alternating compression, and upon the lever arm of the centrifugal force (distance between the roll axis and the center of gravity). Like all rigid bodies, the vehicle body consistently executes a screwing or rotating motion; this motion is supplemented by a lateral displacement along the instantaneous axis.

Figure 15: Roll behavior when cornering
b Track width, h_S Height of center of gravity, R Curve radius, ε Curve banking, G Vehicle weight, F_{FL} Centrifugal force. S Center of gravity, W Roll center.



The higher the roll axis, i.e., the closer it is to a parallel axis through the center of gravity, the lower the vehicle inclination due to cornering with an alternating and parallel spring rate. But because the roll centers of the axes must stray upwards for this purpose, the track change increases. Excessive track changes may give rise to disrupted directional stability when driving on an uneven road and hence poor handling that compromises the feeling of safety. There is also the risk of the body lifting at higher lateral accelerations if the roll centers are too high. Therefore the roll centers are typically under 120 mm.

A frequently applied means for finding the approximate roll axis is based on determining the centers of rotation of an equivalent body motion. This body motion takes place in those two planes through the front and rear axles which are vertical relative to the road. The centers of rotation are those (hypothetical) points in the body which remain stationary during the rotation. The roll axis, in turn, is the line which connects these roll centers (instantaneous centers).

The complexity of the operations required for a more precise definition of the spatial relationships involved in wheel motion make it advisable to employ general, three-dimensional simulation models [1, 2, 3].

The greater the driving speed with the same curve radius, the more closely the driving-dynamic limit range is approached and there is the danger of skidding or overturning (see Table 12).

Table 12: Critical speeds in curves (numerical-value equations)

	Flat curve	Banked curve
Speed at which the vehicle exceeds the limit of adhesion (skid)	$v \le 11.28\sqrt{\mu_r\, r_k}$ km/h	$v \le 11.28\sqrt{\frac{(\mu_r + \tan\varepsilon)\, r_k}{1 - \mu_r \tan\varepsilon}}$ km/h
Speed at which the vehicle overturns	$v \ge 11.28\sqrt{\frac{b}{2}\frac{r_k}{h_s}}$ km/h	$v \ge 11.28\sqrt{\frac{\left(\frac{b}{2\,h_s} + \tan\varepsilon\right) r_k}{1 - \frac{b}{2\,h_s}\tan\varepsilon}}$ km/h

h_S Height of center of gravity (in m), μ_r Max. coefficient of friction, b Track width (in m), r_K Curve radius (in m), ε Curve banking.

Dynamics of vertical motion

Features

The dynamics of vertical motion of a vehicle are determined primarily by its behavior in the vertical direction when driving over road excitations. There are also excitations in the transverse and longitudinal directions and excitations of rotating bodies with imbalance, for example generated by imbalance in the wheel/tire combination, but the vertical excitations of the road are the primary concern here. The frequency range under consideration extends up to approximately 30 Hz. The purpose of the vehicle is to transfer the excitations of the road with the lowest possible accelerations to the vehicle occupants.

The most important parameters for setting up a vehicle are:

- spring and shock-absorber setting, if necessary also active systems such as active control damping,
- kinematics and elastokinematics of the axles,
- tire properties, here above all the spring stiffness in the vertical direction,
- tuning the engine and transmission mounts in both vehicles with internal-combustion engines and electrically powered vehicles.

The dynamics of vertical motion of a vehicle are also referred to as ride comfort; this relates to the haptically perceptible comfort, but not the acoustics.

Road excitation

Before a simplified model of the transfer behavior of a vehicle is presented, the issue of quantifying road excitation will be addressed.

Road excitations are for the purpose of the dynamics of vertical motion divided into two groups:

- irregularity excitations of almost any length,
- individual events, e.g., driving over drain covers.

When describing courses of irregularity in a road of almost any length depending on the path or time, a distinction is made between harmonic (sinusoidal), periodical, and stochastic courses of irregularity (Figure 16).

Harmonic course of irregularity

In this case (Figure 16a), the irregularity height h following distance driven x or point in time t results in:

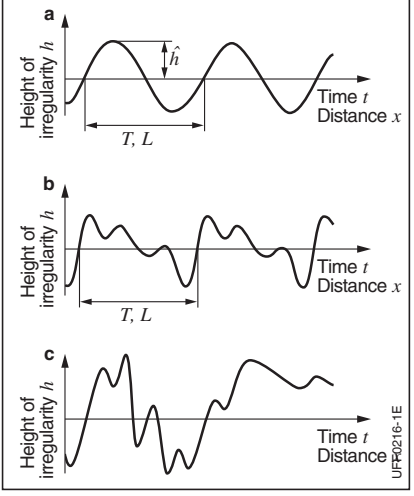
$$h(x) = \hat{h} \sin(\Omega x + \varepsilon), \qquad \text{(Equation 1)}$$

$$h(t) = \hat{h} \sin(\omega t + \varepsilon), \qquad \text{(Equation 2)}$$

where:

$\hat{h}$ Amplitude of irregularity height,
L Period length,
v Speed,
ε Phase displacement,
$\Omega = 2\pi/L$ Breakdown angular frequency,
$\omega = v\,\Omega = 2\pi\, v/L$ Time angular frequency.

Figure 16: Types of courses of irregularity
a) Sinusoidal course of irregularity,
b) Periodical course of irregularity,
c) Stochastic course of irregularity.
T Period duration,
L Period length,
$\hat{h}$ Amplitude of irregularity height.



Periodical course of irregularity
In the next step, it is no longer only purely sinusoidal but also periodical courses of irregularity (Figure 16b) that are assumed, which means that expressing the relationship as a Fourier series (periodical functions can be described as an infinite series of sinusoidal oscillations) results in:

$$h(x) = h_0 + \sum_{k=1}^{\infty} \hat{h}_k \sin(\Omega x + \varepsilon_k), \quad \text{(Equation 3)}$$

$$h(t) = h_0 + \sum_{k=1}^{\infty} \hat{h}_k \sin(\omega t + \varepsilon_k), \quad \text{(Equation 4)}$$

where:
h_0 Base amplitude,
$\hat{h}_k$ Amplitude,
ε_k Phase displacement,
$\Omega = 2\pi/L$ Breakdown angular frequency,
L Period length.

Stochastic course of irregularity
For real road surfaces, i.e., not only for periodical but also for random (stochastic) courses of irregularity (Figure 16c), in the transition into the complex syntax of the summation formula the integral is arrived at, thus moving from the discrete to the continuous spectrum [3]:

$$h(x) = \int_{-\infty}^{\infty} \underline{\hat{h}}(\Omega)\, e^{j\Omega x} d\Omega \quad \text{(Equation 5)}$$

with the continuous amplitude spectrum

$$\underline{\hat{h}}(\Omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} h(x)\, e^{-j\Omega x} dx\,. \quad \text{(Equation 6)}$$

In the time range:

$$h(t) = \int_{-\infty}^{\infty} \underline{\hat{h}}(\omega)\, e^{j\omega t} d\omega \quad \text{(Equation 7)}$$

where

$$\underline{\hat{h}}(\omega) = \frac{1}{v}\, \underline{\hat{h}}(\Omega)\,. \quad \text{(Equation 8)}$$

From the description of the height profile $h(t)$ or $h(x)$ it is possible to obtain power spectral densities $\Phi_h(\Omega)$ (PSD) depending on the breakdown angular frequency. A mathematical definition and detailed discussion of these variables are not provided here. Refer to [1, 3]. For the following examination, the descriptive significance of the power spectral density that exists in the "energy in an infinitesimal frequency band" (with regard to path or time frequency) or the "power output in the natural frequency band" is sufficient.

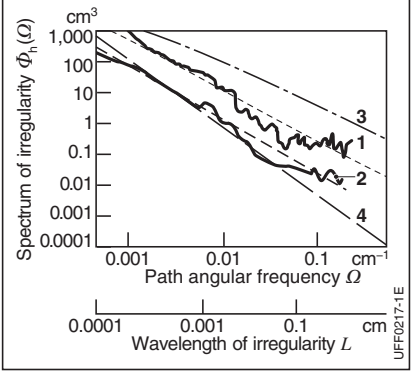
The power spectral densities specify the distribution of the power output of the excitation spectrum within the entire irregularity spectrum. This is usually represented in the double-logarithmic scale (Figure 17). An approximation of the courses of power spectral density on real road surfaces is possible here with straight lines, as described by the following equation:

$$\Phi_h(\Omega) = \Phi_h(\Omega_0) \left(\frac{\Omega}{\Omega_0} \right)^{-w} \quad \text{(Equation 10)}$$

where:
– $\Phi_h(\Omega_0)$ Power spectral density with a reference angular frequency Ω_0 (as a rule $\Omega_0 = 1\ \text{m}^{-1}$, which corresponds to a reference wavelength of $L_0 = 2\pi/\Omega_0 = 6.28$ m),
– w Surface undulation of the road surface (from 1.7 to 3.3; for standard roads $w = 2$).

If the degree of irregularity and the surface undulation of a road surface are known, all the parameters are set in the above linear equation and it can then be used to characterize different types of road surface. The degree of irregularity is defined as the power spectral density at a reference angular frequency, i.e., $\Phi_h(\Omega_0)$.

Figure 17: Irregularity spectrum depending on the breakdown angular frequency.
1 Ordinary road, 2 Expressway/freeway,
3 Upper range limit for poor roads,
4 Lower range limit for good roads.

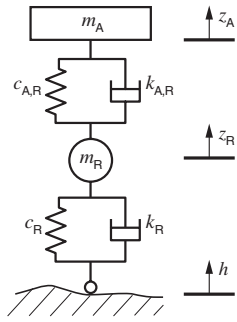


The biggest advantage of power spectral density is that with the ISO standard ISO 8608:2016-11 [9] it sets out a standard so that all stochastic courses of irregularity can be classified from “very good” to “very bad”.

The disadvantage of power spectral density is that individual obstacles, such as for example drain covers, expansion joints on bridges, and so-called sleeping policemen, cannot be correctly resolved even though these are clearly discerned by the driver when driven over in the vehicle. These individual obstacles are therefore considered separately.

Figure 18: Dual-mass oscillator as quarter-vehicle model

m_R Wheel mass, including components that also vibrate (brake, proportional axle mass, etc.),
m_A The body mass supported by a wheel,
c_R Stiffness of the tire springs,
$c_{A,R}$ Wheel-related stiffness of the body springs,
k_R Dampening rate of the tire (usually negligible as $k_R << k_{A,R}$),
$k_{A,R}$ Wheel-related dampening rate of the body damper,
h Vertical excitation,
z_A Body coordinates in vertical direction,
z_R Wheel coordinates in vertical direction.



Quarter-vehicle model

The road excitations take effect on the vehicle body via the tires, wheel suspension, and the suspension/shock-absorber system. For theoretical analyses, vibration models of different complexity can be applied. With rising model complexity, there is an increase in the degrees of freedom. To improve clarity, the fundamental relationships within vibrating vehicle systems will be explained on the basis of the quarter-vehicle model (Figure 18). The name of the model indicates that only one quarter of the entire vehicle is considered. There is therefore only the important degree of freedom in the vertical direction. More detailed information can be found in [1], [3], and [10].

On specification of the masses, the rigidities, and damping factors, all of the parameters have been set in the quarter-vehicle model for technical analyses of the vibrations and two differential equations can be set up with the variables designated in Figure 18:

$$m_A\ddot{z}_A + k_{A,R}(\dot{z}_A - \dot{z}_R) + c_{A,R}(z_A - z_R) = 0\,, \qquad \text{(Equation 11a)}$$

$$m_R\ddot{z}_R - k_{A,R}(\dot{z}_A - \dot{z}_R) - c_{A,R}(z_A - z_R) + c_R z_R + k_R\dot{z}_R = c_R h + k_R\dot{h}. \qquad \text{(Equation 11b)}$$

The division of Equation 11a and Equation 11b by the mass in each case leads to the normal form of a differential equation of the 2nd order and delivers the damped natural frequency ω_g and the undamped natural angular frequency ω_u as well as the damping factors for the wheel D_R and the body D_A.

The following applies to the uncoupled and undamped natural angular frequency at the wheel:

$$\omega_R = \sqrt{\frac{c_R + c_{A,R}}{m_R}} \approx \sqrt{\frac{c_R}{m_R}} \quad \text{(since } c_R \approx 10\ c_A\text{).} \qquad \text{(Equation 12)}$$

The following applies accordingly to the body:

$$\omega_A = \sqrt{\frac{c_{A,R}}{m_A}}\,. \qquad \text{(Equation 13)}$$

In general, the damped natural angular frequency ω_g is calculated according to:

$$\omega_g = \omega_u \sqrt{1 - D^2}\,, \qquad \text{(Equation 14)}$$

whereby the following is also assumed as an approximation:

$$\omega_g \approx 0.9\ \omega_u\,.$$

The following applies to the damping factor D_R at the wheel:

$$D_R = \frac{k_{A,R}}{2m_R\omega_R} = \frac{k_{A,R}}{2\sqrt{(c_R + c_{A,R})\,m_R}} = \frac{m_A\omega_A}{m_R\omega_R}\,D_A\,, \qquad \text{(Equation 15)}$$

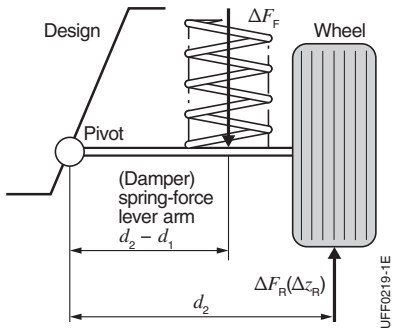
whereby experience has shown that $D_R \approx 0.4$ is to be aimed for. The same applies to the body:

$$D_A = \frac{k_{A,R}}{2m_A\omega_A} = \frac{k_{A,R}}{2\sqrt{(c_{A,R}m_R)}}\,. \qquad \text{(Equation 16)}$$

In this case, $D_A \approx 0.3$ has proven to be effective. The dynamic wheel-load fluctuations ΔG result in:

$$\Delta G = m_R\ddot{z}_R + m_A\ddot{z}_A = c_R(h - z_R) + k_R\,(\dot{h} - \dot{z}_R)\,. \qquad \text{(Equation 17)}$$

Figure 19: Spring ratio
ΔF_F Actual spring force,
ΔF_R Wheel force,
Δz_R Spring compression,
d_2 Distance from wheel center plane to the pivot,
$d_2 - d_1$ Spring-force lever arm.



First design of damping and spring stiffness

A rough design of the suspension/shock-absorber system of a motor vehicle can be made with the results of the quarter-vehicle model.

If a body natural frequency (usually $f_A \approx 1$ to 1.5 Hz) is specified and the body mass (or the proportion of the body mass on a wheel) is known, the wheel-related body spring rigidity can be determined:

$$c_{A,R} = \omega_A^2\ m_A\,. \qquad \text{(Equation 18)}$$

The conversion to the actual rigidity of the body springs takes place taking account of the ratio i between the wheel and spring movement in accordance with Figure 19. First of all, the actual spring force

$$\Delta F_F = c_A \Delta z_F \qquad \text{(Equation 19)}$$

and the wheel force ΔF_R are formulated with a spring compression Δz_R. The following applies to ΔF_R:

$$\Delta F_R = c_{A,R}\,\Delta z_R\,. \qquad \text{(Equation 20)}$$

The torque equilibrium around the pivot in Figure 19 leads to:

$$c_{A.R}\,\Delta z_R\,d_2 = c_A\,\Delta z_F\,(d_2 - d_1)\,. \qquad \text{(Equation 21)}$$

This can be used to convert the actual spring rigidity c_A in accordance with the geometric relationships to the wheel-related rigidity $c_{A,R}$:

$$c_{A,R} = c_A\,i^2 \qquad \text{(Equation 22)}$$

with the spring ratio i

$$i = \frac{(d_2 - d_1)}{d_2} = \frac{\Delta z_F}{\Delta z_R}\,. \qquad \text{(Equation 23)}$$

The same holds true for the vibration damper. For calculation of the body dampening rate (its effect on the wheel), the following applies in relation to the wheel in accordance with Equation 16:

$$k_{A,R} = 2\,D_A\,\sqrt{c_{A,R}\,m_R}\,. \qquad \text{(Equation 24)}$$

With $D_A = 0.3$ (see above) and m_A as a known variable of the vehicle under examination, the body dampening rate can be determined taking account of Equation 23.

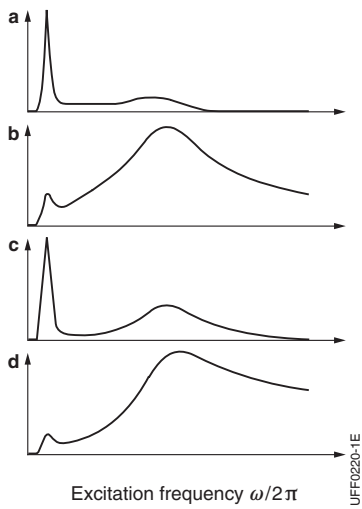
Conflict of aims in the first design of body spring stiffness and body damping

Equation 11a and Equation 11b can be used to calculate the transfer functions (enlargement functions). The amount of the transfer function is the quotient from the stationary proportion of the amplitude of the output variables (body acceleration or wheel-load fluctuation) and the amplitude of the input variable (vertical excitation, see Figure 18). The principle of the progression is shown in Figures 20c and 20d.

These transfer functions are now linked to the road irregularities. The results are the power spectral densities for the body acceleration (Figure 20a) and for the wheel-load fluctuation (Figure 20b) [3].

Figure 20: Transfer function and power spectral density for body acceleration and wheel-load fluctuations

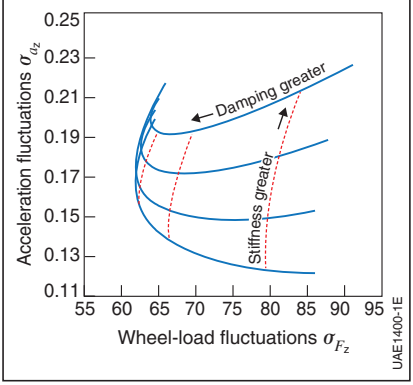
(According to [3])
a) Power spectral density, body acceleration,
b) Power spectral density, wheel-load fluctuation,
c) Amount of transfer function, body acceleration,
d) Amount of transfer function, wheel-load fluctuation.



The higher excitation amplitudes of the road surface at low frequencies raise the vibration amplitudes in the natural frequency range of the body (1 to 2 Hz). In the natural frequency range of the wheel (10 to 14 Hz), however, the low irregularity amplitudes lead to drops in the power spectral densities compared to the transfer functions, which means the body movements become more dominant than the axle movements.

From the power spectral densities of the body accelerations and the wheel loads it is possible to calculate in each case the fluctuation of the body acceleration and the wheel load as a standard deviation [3]. When these two quantities for different body-stiffness values and body-damping values are plotted (Figure 21), it can be seen that it is possible for example with slight damping to reduce the fluctuations of the body acceleration and thereby improve ride comfort. At the same time, however, the wheel-load fluctuations increase and the situation may arise where the tire is no longer in contract with the ground, potentially resulting in wheel hopping. In a wheel-hopping situation there is the possibility of the tire generating lateral forces and consequently the vehicle potentially tramping, i.e., vehicle safety is diminished.

Figure 21: Relationship between wheel-load fluctuations and body-acceleration fluctuations as a function of body stiffness and body damping



Realistic vehicle models for the dynamics of vertical motion

Complete-vehicle models based on general multiple-body simulation programs are used to predict the dynamics of vertical motion. In this way, the component properties are directly parameterized and do not have to be converted. In addition, all the degrees of freedom of the body are then naturally excited, i.e., aside from the vertical direction, as in the quarter-vehicle model, lateral and longitudinal motion.

The torsion bars on the front and rear axles can also be evaluated with a complete-vehicle model. In this way, the vibrational behavior of the wheel and the engine/transmission combination can also be analyzed in order to come up with a favorable design of the engine and transmission mounts with the body damping and body stiffness.

For further refine the dynamics of vertical motion the body damping and body stiffness not only are defined by constant values as in the quarter vehicle but have nonlinear identifiers.

In the case of real vehicle axles the so-called secondary spring rate must be taken into account as well as the friction. The secondary spring rate can, depending on the axle type, be up to 20 % of the spring rate of the body spring. The reason for this is that the rubber mounts installed in the axle also possess a stiffness which becomes noticeable when the axle is deflected. Thus, the expressions found for the stiffness of the body spring and damping of the body damper are initial guide values.

Objectification of subjective perception

Only a short time after the introduction of the automobile, the first approaches to objectifying ride comfort were developed at the start of the 20th century. The influence of comfort – particularly in the case of professional drivers – on fitness (i.e., ensuring good physical and mental condition) as well as on the health of the driver and thus on overall road safety was recognized very quickly. Accordingly, attempts were made to shape meaningful variables to describe comfort and to compare these in the form of objective comfort factors for different variants of vehicles and roads. Since then, a number of different assessment methods have been defined, based almost exclusively on acceleration signals at and around the driver's seat. Here, the frequency-dependent sensitivity of people to vibrations and oscillations initiated at different positions as well as different types and directions (translatory or rotatory; x, y, z direction) is taken into account with the help of the corresponding frequency weighting filters.

Two widely used approaches are the VDI guideline 2057 [11] and the standard ISO 2631 [12]. Despite the relatively low overhead involved in these analyses compared to other methods, these have good informative value. At the same time, these form the basis for many approaches to comfort assessment that were developed in the 1990s and later (see [11], [13], [14], [15], [16]). In the methods used to date, it is normally the case that the accelerations caused by the excitation at different measuring points (initiation positions) are used to determine an overall comfort factor (possibly from a number of partial comfort factors at different measuring points) that is to represent a measure of the load on the occupants. The accelerations are first of all frequency-weighted (i.e., amplitudes of different frequencies are linked to frequency-dependent factors) and subsequently, for example with the RMS method (Root Mean Square, see [11]) in accordance with:

$$\overline{a}_{\mathrm{RMS}} = \sqrt{\frac{1}{T}\int_0^T a_{\mathrm{w}}^2(t)\,dt} \qquad \text{(Equation 26)}$$

or with the RMQ method (Root Mean Quad, see [14]) in accordance with:

$$\overline{a}_{\mathrm{RMQ}} = \sqrt[4]{\frac{1}{T}\int_0^T a_{\mathrm{w}}^4(t)\,dt} \qquad \text{(Equation 27)}$$

reduced to a value. Depending on the number of evaluated measuring points, an overall factor has to be formed in a subsequent operation. For the VDI guideline 2057, only the evaluation of the measuring point with the highest load is relevant. All translatory accelerations of this initiation position (hand, seat or foot) are taken into account equally in the calculation of the factor. However, no uniform specifications for weighting individual initiation positions have yet been developed.

Only in the recent past have the results of Weber, Fechner, and Stevens from the field of psychophysics [17] been taken into account in the objectification of subjective perception. Thus, the following effects must be taken into account in the evaluation of objective measurement data:

- lower perceptibility threshold,
- difference perceptibility threshold between two signals,
- order in which the signals are evaluated (order effect),
- masking effect.

Evaluation methods which take the Weber-Fechner law into account are much more accurate than simple absolute evaluations [18]. Nevertheless, these evaluation methods must be supported by a real road test or by trials with test subjects on suitable simulators.

Special operating dynamics for commercial vehicles

Essentially, commercial vehicles are governed with regard to driving dynamics by the same laws as for passenger cars. Supplementary aspects are described in the following. Typical differences between passenger cars and commercial vehicles are:

– The typical unladen curb weight and the gross weight rating for commercial vehicles is higher than for passenger cars. In Europe a typical heavy truck with two axles has a gross weight rating of 18 t; a typical semitrailer unit with a total of five axles has a gross weight rating of 40 t.
– Heavy trucks and buses often have more than two axles so as to be able to achieve greater vehicle weights. The difference between the vehicle weight in the laden and unladen states is much greater in trucks and light utility vans than in passenger cars.
– Trucks are very often operated with trailers.

Dynamics of linear motion of commercial vehicles

As in passenger cars the rolling resistance of a commercial vehicle comprises rolling resistance, aerodynamic drag, and if necessary climbing resistance. Because of the high weight, for heavy trucks on gradients greater than approx. 1 % climbing resistance is the dominant contributor to rolling resistance [19].

The permitted top speed of max. 80 km/h or 90 km/h for heavy commercial vehicles in the majority of European countries means that aerodynamic drag – unlike for passenger cars – never becomes the dominant factor.

Because commercial vehicles are limited in length in the EU (and many other countries), the front of the vehicle is essentially a vertical surface. The achievable c_w value is accordingly much higher than the value for passenger cars. An aerodynamically sound, modern truck used in long-distance haulage has a c_w value of around 0.5. For special superstructures much higher values can be expected (e.g., concrete mixers, timber trucks). The frontal area of a long-distance truck, on account of the legal requirements for width and height, amounts to around 10 m^2 (4 m high by 2.5 m wide).

Dynamics of lateral motion of commercial vehicles

Chassis system configuration

The aim is to achieve non-problematic, understeering handling. Stationary vehicle handling is defined as for passenger cars by the self-steering effect. It is influenced by

– the axle configuration,
– the geometry of the complete vehicle (wheelbase, track width),
– the axle-load distribution,
– the tires,
– the mechanical and hydraulic parameters of the steering gear,
– the stiffness and geometry of the steering linkage,
– the elastokinematics of the front and rear axles,
– the frame stiffness, and
– the roll stabilization at the front and/or rear axles.

The self-steering effect of leaf-spring axles (“steel suspension”) is essentially determined by the inclination of the leaf springs in the longitudinal direction of the vehicle. Because the inclination of the leaf-spring branches changes with the load, the self-steering effect of leaf-spring axles is dependent on the vehicle load. This effect is referred to as “roll steer” of the axle in that it depends on the roll angle of the vehicle when cornering as well as on the compression/jounce.

Steering movements that result from the elasticity of axle bearings are a further component of self-steering. These are dependent on the cornering force on the wheel and are therefore referred to as “lateral-force steer”.

Heavy-duty trucks with air suspension are normally equipped with solid or rigid axles controlled with suspension arms and links. Self-steering can be influenced by a specific design of the flexibility of the link bearings. Because these axles are normally also fitted with ride-height control, the kinematic position of the links

and thus the "roll steer", in contrast to leaf-spring axles, are not dependent on the vehicle load.

External interference factors such as road-surface irregularities or crosswind, for example, must not give rise to any great disruptions of vehicle motion. The elastokinematic design of the chassis serves to minimize these interference factors. On the front axles the aim is to minimize steering movements as a result of parallel and alternating compression and rebound and as a result of braking. The kinematic points of the steering, for example, act as the setting levers in the configuration of the elastokinematics.

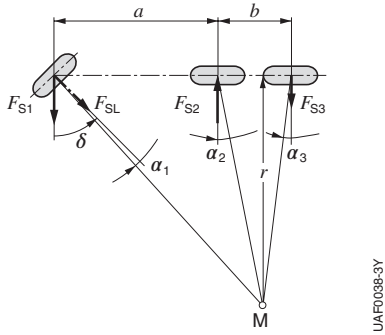
In the case of leaf-spring axles the connecting points of the leaf springs and the leaf-thickness progressions are important setting levers.

Geometrically nonlinear finite-element programs are used in the calculation to configure the elastokinematics. The configuration is monitored on elastokinematic test benches and in the complete-vehicle road test.

To date, axle control concepts involving independent wheel control have only been implemented on light utility vans and buses.

Figure 22: Cornering forces F_S and slip angle α on a three-axle vehicle with non-steered tandem axle

a Distance, front axle to first rear axle,
b Distance, first rear axle to second rear axle,
F_{Si} Cornering forces, i = 1, 2, 3, L,
α_i Slip angle, i = 1, 2, 3,
δ Steering angle,
r Radius,
M Instantaneous center.



In vehicles with multiple non-steered rear axles, such as 6×4 or 6×2 (six wheels, four or two of which are driven), a constraining torque is generated around the vehicle's vertical axis on account of the different slip angle at the first and second rear axles. The resulting lateral force on the axles is relevant for the maneuvering system and can be determined for small angles α ($\tan\alpha \approx \alpha$ in the radian measure; error < 1 % for α < 10 ° or p < 17 %) as follows (Figure 22; quantities, see key to Figure 22):

$F_{S1} = F_{S2} - F_{S3}$ and
$F_{S1}\,a = F_{S3}\,b$ (moment equilibrium),
where
$F_{S2} = c_{p2}\,n_2\,\alpha_2$,
$F_{S3} = c_{p3}\,n_3\,\alpha_3$.

Slip angle:

$$\alpha_2 = \frac{1}{r} \cdot \frac{c_{p3}\,n_3\,b\,(a+b)}{c_{p3}\,n_3(a+b) + c_{p2}\,n_2\,a},$$

$$\alpha_3 = \frac{b}{r} - \alpha_2.$$

– α_i is obtained in the radian measure,
– c_{p2} and c_{p3} are the slip stiffnesses from the tire maps,
– n_2 and n_3 is the number of tires per axle.

Tipping resistance

As the vehicle's total height increases, the vehicle will also have an increasing tendency to tip to the side in a turn before it starts to slide. In the same way as the overall dynamic behavior, the tipping limits of vehicles are determined by means of multibody-system simulations (multibody system, Figure 23). The simulations investigate various stationary and non-stationary maneuvers, e.g., steady-state circular-course driving ([20]) and double lane change [21]. The achievable lateral acceleration b at the tipping limit during steady-state circular-course driving is b = 6…8 m/s² for light utility vans and b = 4…6 m/s2 for trucks and buses.

The driving-dynamics control systems prescribed by law (Electronic Stability Program) predominantly used in today's commercial vehicles have functions which reduce the risk of tipping. The interven-

tion threshold of these systems can be adapted within limits to the load and load distribution of the vehicle whereby these variables are estimated by an algorithm from the vehicle's instantaneous handling.

Width requirement
The width requirement of motor vehicles and truck-trailer combinations is greater during cornering than when the vehicle moves in a straight line. With respect to selected driving maneuvers, it is necessary to determine the radius described by the vehicle's outer extremities during cornering, both in order to ascertain its suitability for certain applications (e.g., narrow transit routes through constricted areas) and to confirm compliance with legal requirements [22] (turning-circle regulation, Figure 24). This is determined using multibody-system simulations or by means of real tests.

Handling characteristics
Objective analyses of vehicle handling are based on various driving maneuvers (see Driving-dynamics test procedures). For example, the vehicle's steady-state self-steering properties and roll behavior are assessed with the "Steady-state circular-course driving" test procedure. Moderately understeering self-steering properties that are easily manageable by the driver in this test procedure also provide

Figure 24: Turning-circle regulation [22] on example of a semitrailer unit
Representation of the permissible steady-state cornering radii.

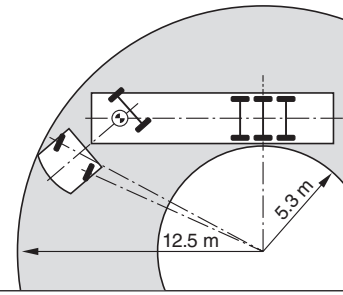
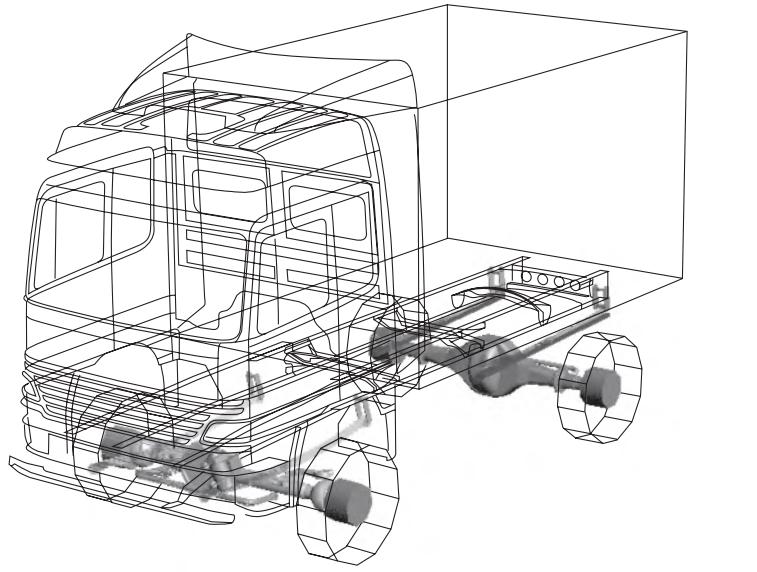


Figure 23: Multibody-system model of an all-wheel drive vehicle
for determining self-steering effects and simulating operating dynamics.

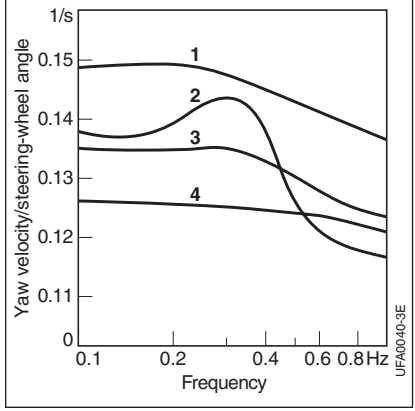


a significant basis for a good subjective driving impression in dynamic steering maneuvers and in braking maneuvers. Dynamic handling can be evaluated for example with the “sinusoidal steering input” maneuver. Test procedures are additional used to evaluate driving stability in braking operations (e.g., braking in a turn/curve) and to evaluate the steering “feel” (“weave test”).

The dynamic lateral response of truck-trailer combinations (e.g., semitrailer, truck-trailer) generally differs from that of rigid vehicles. Particularly significant are the distribution of loads between truck and trailer or semitrailer, and the design and geometry of the mechanical coupling device within a given combination.

The vehicle's straight-running stability can be impaired by a variety of factors. These include gusts of crosswind, uneven road surfaces, obstacles on one side, ruts in the road surface, and crossfall cambers. The yaw oscillations caused by these impairments must decay rapidly if vehicle stability is to be maintained. These yaw oscillations can be assessed on the basis of the yaw-velocity frequency response (Figure 25). The yaw-velocity frequency responses of different truck-trailer combinations demonstrate excessive increase in resonance in the least favorable case (tractor-trailer unladen, center-axle trailer laden, curve 2 in Figure 25). This type of combination demands a high degree of driver skill and a defensive driving style.

Figure 25: Yaw-velocity frequency responses
1 Semitrailer unit (laden),
2 Truck-trailer unit (unladen with laden center-axle trailer),
3 Truck-trailer unit (laden), 4 Truck (laden).



With semitrailer units, braking maneuvers undertaken under extreme conditions can induce jackknifing. This process is initiated when, on a slippery road surface (μ-low), loss of lateral force is induced by excess braking force applied to the tractor’s rear axle or due to excess yaw moment under μ-split conditions. Installation of an driving-dynamics control system (Electronic Stability Program) represents the most effective means of preventing jackknifing.

References

[1] M. Ersoy, S. Gies: Fahrwerkhandbuch, 5th Edition, Verlag Springer Vieweg, 2017.
[2] H.-P. Willumeit: Modelle und Modellierungsverfahren in der Fahrzeugdynamik, Teubner-Verlag, 1998.
[3] M. Mitschke, H. Wallentowitz: Dynamik der Kraftfahrzeuge, 5th Edition, Springer-Verlag 2014.
[4] ÖNORM V 5050: Traffic accident and vehicle damage – Terminology.
[5] Fritz Sacher in Fucik, Hartl, Schlosser, Wielke (Editors): Handbuch des Verkehrsunfalls, Part 2, Manz-Verlag Vienna, 2008.
[6] 71/320/EEC: Council Directive of 26 July 1971 on the approximation of the laws of the Member States relating to the braking devices of certain categories of motor vehicles and their trailers.
[7] C. Kobetz: Modellbasierte Fahrdynamikanalyse durch ein an Fahrmanövern parameteridentifiziertes querdynamisches Simulationsmodell, Dissertation, TU Wien (Vienna), 2003.
[8] T. Schütz: Hucho-Aerodynamik des Automobils, 6th Edition, Verlag Springer Vieweg, 2013.
[9] ISO 8608:2016-11 Mechanical vibration – Road surface profiles – Reporting measured data; published by the International Organization of Standardization.
[10] J. Reimpell, J.W. Betzler: Fahrwerktechnik – Grundlagen. 5th Edition, Vogel Verlag, 2005.
[11] VDI Directive 2057: Human exposure to mechanical vibration, whole-body vibration.
VDI-Gesellschaft Entwicklung Konstruktion Vertrieb, Düsseldorf 2002.
[12] ISO 2631 AMD1 (2010): Mechanical vibration and shock – Evaluation of human exposure to whole-body vibration.
[13] S. Cucuz: Auswirkung von stochastischen Unebenheiten und Einzelhindernissen der realen Fahrbahn. Dissertation, TU Braunschweig, 1992.
[14] D. Hennecke: Zur Bewertung des Schwingungskomforts von Pkw bei instationären Anregungen. VDI Progress Reports, Düsseldorf, 1995.
[15] M. Mitschke, S. Cucuz, D. Hennecke: Bewertung und Summationsmechanismen von ungleichmäßig regellosen Schwingungen. ATZ 97/11 (1995).
[16] Kudritzki: ridemeter-calculated ride comfort, SAE 2007-01-2388.
[17] Weber-Fechner law in https://de.wikipedia.org/wiki/Weber-Fechner-Gesetz
[18] B. Jörissen: Objektivierung der menschlichen Schwingungswahrnehmung unter Einfluß realer Fahrbahnanregungen, Dissertation, University of Duisburg-Essen, 2011.
[19] M. Hilgers: Getriebe und Antriebsstrangauslegung. Verlag Springer Vieweg, 2016.
[20] ISO 14792 (2011): Road vehicles – Heavy commercial vehicles and buses – Steady-state circular tests.
[21] ISO 3888-1 (1999), ISO 3888-2 (2011): Passenger cars – Test track for a severe lane-change manoeuvre –
Part 1: Double-lane change
Part 2: Obstacle avoidance
[22] Commission Directive 2003/19/EC of 21 March 2003 for changing the Guideline 97/27/EC of the European Parliament and of the Council about the weights and dimensions of certain classes of motor vehicles and motor-vehicle trailers regarding the adaptation to the technical progress.

Driving-dynamics test procedures

Assessment of vehicle handling

Overall system

The science devoted to studying the dynamics of vehicle handling generally defines its subject as the overall behavior of the entire system represented by "driver – vehicle – environment" (Figure 2). The driver assesses the vehicle's handling qualities based on the sum total of his/her subjective impressions. Internationally standardized test procedures have been developed since the end of the 1970s. On the one hand, these test procedures are intended to describe vehicle handling as objectively and consistently as possible. On the other hand, they serve to determine characteristic variables which correlate well with the subjective driving impressions.

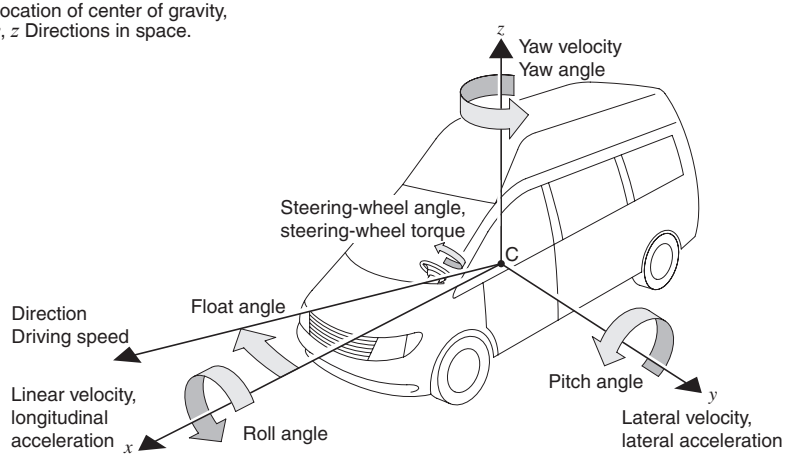
Most of the internationally standardized procedures are carried out in an open loop, i.e., with a defined steering input without controlling intervention by the driver. In this way, the result is not influenced by different steering strategies of individual drivers. Only a few procedures are carried out in a closed loop with the input of a specific course which can be taken by the driver in an individual way.

There are currently a total of roughly 20 ISO standards for driving-dynamics procedures, of which the following maneuvers used most frequently in vehicle development will be described here:
- Steady-state circular-course driving [1], [2].
- Test procedures for transient response [3], [4].
- Weave test and transition test [5], [6].
- Braking in a turn/curve [7], [8].
- Closing curve [9].

Further test procedures not explicitly described here are, for example:
- Double lane change (closed loop, [10]).
- Lateral stability of vehicle combinations [11], [12].
- Crosswind stability [13].
- Load changes when cornering [14].

These test procedures were originally created for automobiles ([15], [16]). Test methods based on these standards were later developed for heavy-duty commercial vehicles in order to cater for the unique handling characteristics of heavy-

Figure 1: Assessment quantities for driving dynamics
S Location of center of gravity,
x, y, z Directions in space.



duty commercial vehicles with their large mass and inertia figures.

General boundary conditions which must be equally observed for all driving-dynamics test procedures, for example, the condition of the road surface, environmental conditions and tires, are defined in a separate ISO standard (ISO 15037, [17]), which additionally defines what is required of the measuring techniques for driving-dynamics measurements.

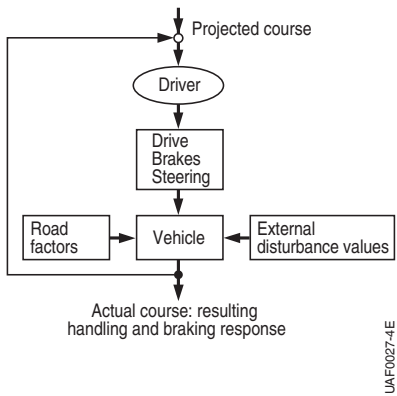
Assessment quantities

The following measurable quantities are primarily used to assess vehicle handling (Figure 1):

- Steering-wheel angle and steering-wheel torque.
- Lateral acceleration.
- Yaw velocity.
- Roll angle.
- Float angle.

Depending on the test procedure defined, characteristic values are determined from these and other quantities for the purpose of describing and assessing vehicle handling. These measured values and characteristic quantities are defined in a separate ISO standard (ISO 8855 [18]).

Figure 2: Overall system "Driver – Vehicle – Environment" as closed loop



Driving-dynamics test procedures as per ISO

Steady-state circular-course driving

Performance

The "Steady-state circular-course driving" test procedure ([1], [2]) is generally carried out in such a way that the vehicle is accelerated on a circular course with a constant radius (a radius of 40 m is usually selected for passenger cars, 80 m for trucks) starting from a very slow starting speed up to the maximum attainable vehicle lateral acceleration. The longitudinal acceleration here should be significantly below 1 m/s^2 so that the driving state can be viewed as sufficiently steady-state. When semitrailer-truck or truck-trailer combinations are measured, the measurement is conducted at a number of constant driving speeds on a specified circular-course radius.

Assessment

The assessment criteria primarily used are the curves of steering-wheel angle, roll angle and float angle plotted against lateral acceleration, which are shown in Figure 3 for a passenger car and for a truck with semitrailer.

From the curve of steering-wheel angle against lateral acceleration it is possible to determine the vehicle's self-steering effect using a single-lane model (see Dynamics of lateral motion). Today's vehicles, both passenger cars and heavy-duty commercial vehicles, are basically designed as understeering. In other words, as their speed increases on a constant circular-course radius, they require a significant increase in steering angle. The limit range of an understeering vehicle is determined by the maximum transferable cornering force of the front wheels, which is recognizable from the sharply progressively rising steering-wheel angle with high lateral acceleration. It is clear from the representation that as a rule higher lateral accelerations are possible for automobiles than for trucks.

The curve of roll angle against lateral acceleration describes the lateral inclination of the vehicle that can be perceived by the driver. This is greatly dependent on the vehicle load, particularly in the case

of commercial vehicles. In the case of commercial vehicles with a high loading center of gravity, the maximum achievable lateral acceleration is for the most part determined not by the cornering force of the tires, but by the tipping limit of the vehicle.

In the case of trucks, the roll angle is additionally determined not only at the center of vehicle gravity, but also, due to the torsionally weak frame and the separate driver's cab mounting, at several fixed measuring points.

The curve of float angle against lateral acceleration is an indicator of the lateral stability of the vehicle that can be perceived by the driver. It is determined above all by the properties of the tires. In addition to the tire properties, the following vehicle parameters have a significant influence on stationary vehicle handling:
– Load and axle-load distribution.
– Stiffness of springs and stabilizers.
– Kinematic and elastokinematics movements of the wheel suspension.

Transient response

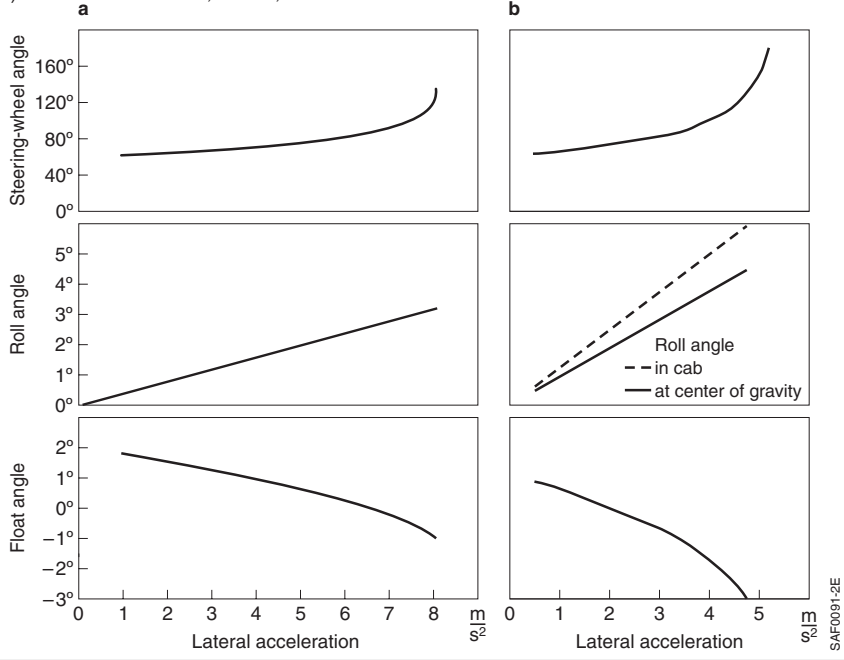
The test procedures for transient response ([3], [4]) serve to determine the vehicle response to rapid, dynamic steering stimulations, e.g.,during rapid evasion maneuvers. Procedures frequently used in tests and simulation are "Step input" and "Sinusoidal steering input and frequency response".

Step input

Performance

The "Step input" test procedure involves the vehicle, starting from straight-ahead driving at a constant speed, being turned in with a very fast steering movement from the zero steering-wheel position to a stationary steering-wheel angle. This results after the transient condition in circular driving with a defined, constant lateral acceleration. An arrangement usual for passenger cars involves turning in with a steering-wheel angle velocity of approximately 360°/s to a lateral acceleration

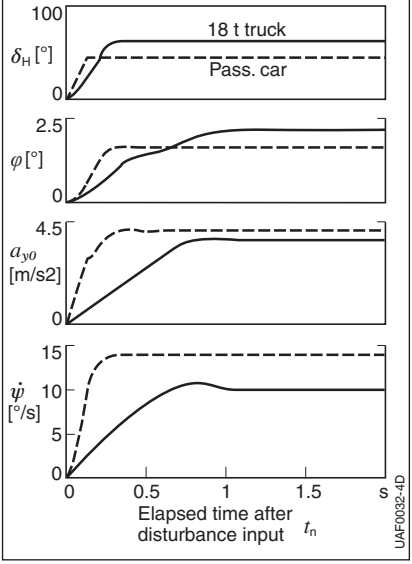
Figure 3: Steady-state circular-course driving
a) Passenger car, 40 m circle radius,
b) 40-ton semitrailer unit, loaded, 80 m circle radius.



of 4 m/s² at a driving speed of 80 km/h. Trucks are often driven at a speed of 60 km/h (Figure 4).

Assessment
The time delay and the overswing values with which the quantities of yaw velocity, lateral acceleration, roll angle, and float angle respond to the steering-wheel angle input are determined during this maneuver (Figure 4). The vehicle should, on the one hand, not respond too sluggishly to the steering movement and, on the other hand, not respond with excessive overswing to the steering input. Figure 4 shows that a truck reacts much more sluggishly than a passenger car.

Figure 4: Step input
(Unladen, driving speed 60 km/h.)
δ_H Steering-wheel angle,
φ Roll angle,
a_{y0} Lateral acceleration,
$\dot{\psi}$ Yaw velocity.

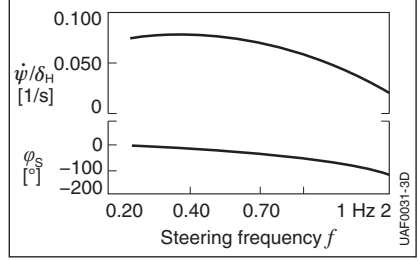


Sinusoidal steering input and frequency response
Performance
The "Sinusoidal steering input and frequency response" test procedure involves the vehicle being excited at a constant driving speed – for example 80 km/h – with a sinusoidal steering-wheel angle signal whose amplitude is kept constant and whose frequency is increased starting at slow steering movements of 0.2 Hz to fast steering movements up to 2.0 Hz. This makes it possible to assess the dynamic vehicle handling in the entire steering-frequency range to be excited by the driver.

The steering amplitude is kept constant during the test and is as a rule selected in such a way that the vehicle remains in the linear driving range, i.e., in the case of passenger cars in such a way that a lateral acceleration of max. 4 m/s² is obtained at the slowest steering frequency. In the case of heavy commercial vehicles, lower maximum lateral accelerations are selected for the most part.

Assessment
For assessment, the amplitudes referred to the input variable of steering-wheel angle and phase angles of the quantities of steering-wheel torque, yaw velocity, lateral acceleration, roll angle, and float angle are determined and plotted against steering frequency. Figure 5 shows such an assessment for an 18-ton truck using the example of the frequency response of

Figure 5: Sinusoidal steering-angle input
(18-ton truck, driving speed 60 km/h.)
δ_H Steering-wheel angle,
$\dot{\psi}$ Yaw velocity,
φ_s Phase angle,
δ Amplitude of steering-angle input.



the yaw velocity. The position of the natural frequencies, the banking that occurs and the phase angles can be used as an assessment criterion for vehicle agility and stability in the case of dynamic steering excitations. Heavy banking in the curve of the yaw-velocity amplitude would for example indicate "nervous" vehicle handling at a specific steering frequency. Large phase angles and a sharply falling amplitude curve indicates on the other hand a very sluggish response by the vehicle in the case of fast steering movements.

As well as the vehicle parameters which influence the stationary handling, the damping properties and the moments of inertia of the vehicle as well as the dynamic properties of tires, steering system and wheel suspension are crucial to vehicle handling in the case of dynamic driving maneuvers.

Weave test and transition test

The "Weave test" and "Transition test" assessment procedures ([5], [6]) were developed to develop the vehicle's response to small, slow steering movements about the steering zero position. The behavior of the vehicle and steering system during the maneuvers correlates well with the driver's subjective impression of the monitorability of the vehicle in everyday driving.

Transition test

Performance

In the transition test the vehicle is turned in, starting from straight-ahead driving at a constant driving speed, with a slow steering movement (e.g., at 80 km/h with a steering-wheel angle velocity of 5 °/s) to a circular course with a low lateral acceleration of 1 to 2 m/s^2. In practice this corresponds to the operation of turning at low speed into a curve on a secondary road.

Assessment

Above all, the time characteristics of the quantities of yaw velocity and steering-wheel torque are assessed. Among others, the time after which the vehicle responds with a quantity of yaw velocity that can be perceived by the driver to the steering excitation is a suitable assessment quantity.

Weave test

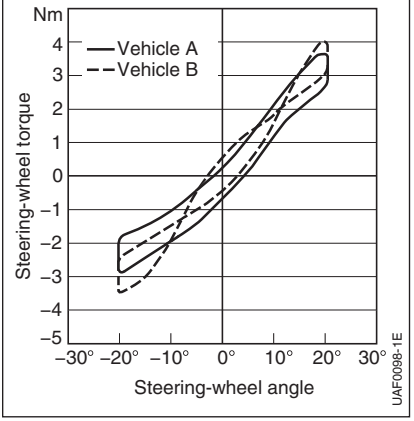
Performance

In the weave test the vehicle is excited at a constant driving speed (e.g., 80 km/h) with a sinusoidal steering excitation at a slow steering frequency of 0.1 to 0.2 Hz. The maximum lateral acceleration here ranges between 2 m/s^2 (usual for trucks) and 4 m/s^2 (usual for passenger cars), i.e., the vehicle is constantly in the linear driving range.

Assessment

To assess the vehicle handling and steering behavior, the curves for the quantities of steering-wheel torque, yaw velocity, lateral acceleration, and float angle as a function of the input steering-wheel angle are used. The non-linearities (among others friction) in the steering system, tires and wheel suspension give rise to hysteresis loops, like the hysteresis loop of steering-wheel torque vs. steering-wheel angle shown by way of example in Figure 6. The ranges of the hystereses and the rise gradients of the curves are suitable assessment criteria. For example, the steering behavior of a vehicle which demonstrates on overlapping hysteresis loop, like vehicle B in Figure 6, is perceived by the driver to be "indifferent about the center position" and "poor in straight-running stability".

Figure 6: Steering hysteresis in weave test



Braking in a turn/curve

Performance

The "Braking in a turn/curve" test procedure ([7], [8]) is likewise geared towards a situation that occurs frequently in real driving operation. Starting out from stationary cornering on a circle with a constant radius with a lateral acceleration of between 3 m/s^2 (usual value for trucks) and 5 m/s^2 (usual value for passenger cars), the vehicle is braked with a constant steering-wheel angle and a defined deceleration, set by the brake-pedal travel required for this purpose. Several tests are carried out in which the brake-pedal setting and thus the braking deceleration are varied from light braking through to full braking.

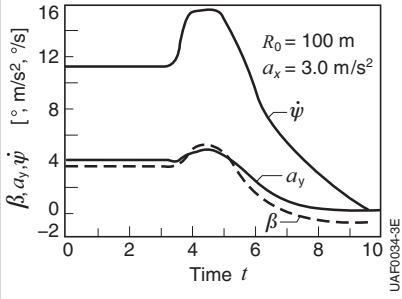
Assessment

Figure 7 shows the time characteristics of yaw velocity, lateral acceleration and float angle of a 7.5-ton truck during a braking operation with a deceleration of 3 m/s^2. Once the braking has started (time 3.0 s), a clear banking of the driving-dynamics quantities can be made out. It shows that the vehicle turns in to the inside of the curve during the braking operation.

The extent of the banking of the driving-dynamics quantities, which are usually plotted against the varied longitudinal deceleration, are used as an assessment criterion for vehicle stability [19].

Figure 7: Time curves for braking in a turn/curve (7.5-ton truck, laden)

Vehicles runs a steady-state circular course. Braking begins at time of 3 s.
$\dot{\psi}$ Yaw velocity, β Float angle,
a_y Lateral acceleration,
a_x Longitudinal deceleration, R_0 Circle radius.



As well as the vehicle parameters which influence vehicle handling in the driving maneuvers described up to now, the braking-force distribution between front and rear axles and the design of the brake-control systems ABS and driving-dynamics control (Electronic Stability Control, ESC) have a significant influence on vehicle handling during this maneuver. It is therefore used among other things to tune the brake system and brake-control systems.

Closing curve

The "Closing Curve" test procedure [9] was developed against the backdrop of the change to the ECE regulation ECE-R13, which requires that as of 2014 virtually all heavy commercial vehicles be duly equipped with ESC systems.

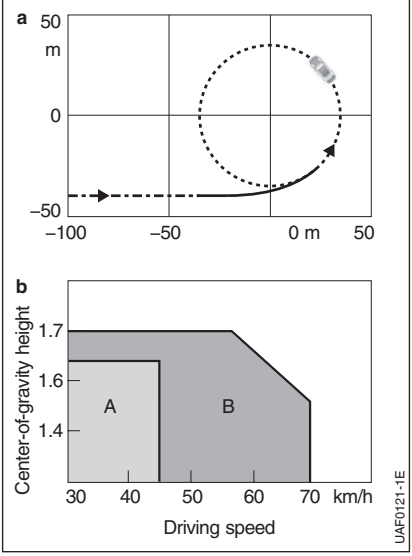
Performance

This test procedure too is based on a real driving situation, namely negotiating an expressway entrance or exit at too high a driving speed. The enhanced safety afforded by the ESC control system can be gaged with this test. To this end, a trajectory leading rapidly from straight-ahead driving into a curve with a tight curve radius is defined. The course progression is created in such a way that for a certain initial speed a defined increase in the lateral acceleration is obtained, e.g., 2 m/s^2 (Figure 8a).

Assessment

The test procedure is carried out at different speeds and center-of-gravity heights, in each case with ESC activated and deactivated. The lift of the inner cornering wheels is used to assess whether the maneuver is classed as "passed". Figure 8b shows, as a function of center-of-gravity height and driving speed, the areas in which the maneuver has passed. The enhanced safety provided by ESC becomes visible.

Figure 8: Enhanced safety due to ESC in the "Closing Curve" test procedure
a) Test progression:
– straight-ahead driving before the circle,
– entry into the circle,
– circular course with constant radius.
b) Areas with stable vehicle response, i.e., safely run course progression.
A ESC deactivated,
B ESC activated.



References

[1] ISO 4138 (2012): Passenger Cars – Steady-state circular driving behaviour – Open-loop test methods.
[2] ISO 14792 (2011): Road vehicles – Heavy commercial vehicles and buses – Steady-state circular tests.
[3] ISO 7401 (2011): Road vehicles – Lateral transient response test methods – Open-loop test methods.
[4] ISO 14793 (2011): Road vehicles – Heavy commercial vehicles and buses – Lateral transient response test methods.
[5] ISO 13674: Road vehicles – Test method for the quantification of on-centre handling –
Part 1: Weave test (2010),
Part 2: Transition test (2016).
[6] ISO 11012 (2009): Heavy commercial vehicles and buses – Open-loop test methods for the quantification of on-centre handling – Weave test and transition test.
[7] ISO 7975 (2019): Passenger cars – Braking in a turn – Open-loop test method.
[8] ISO 14794 (2011): Heavy commercial vehicles and buses – Braking in a turn – Open-loop test methods.
[9] ISO 11026 (2010): Heavy commercial vehicles and buses – Test method for roll stability – Closing-curve test.
[10] ISO 3888: Passenger cars – Test track for a severe lane-change manoeuvre –
Part 1: Double lane-change (2018),
Part 2: Obstacle avoidance (2011).
[11] ISO 9815 (2010): Road vehicles – Passenger-car and trailer combinations – Lateral stability test.
[12] ISO 14791 (2014): Road vehicles – Heavy commercial vehicle combinations and articulated buses – Lateral stability test methods.
[13] ISO 12021 (2010): Road vehicles – Sensitivity to lateral wind – Open-loop test method using wind generator input.
[14] ISO 9816 (2018): Passenger cars – Power-off reaction of a vehicle in a turn – Open-loop test method.
[15] A. Zomotor: Fahrwerktechnik – Fahrverhalten. 2nd Ed., Vogel-Verlag, 1991.
[16] M. Mitschke, H. Wallentowitz: Dynamik der Kraftfahrzeuge. 5th Ed., Verlag Springer Vieweg, 2014.
[17] ISO 15037: Road vehicles – Vehicle dynamics test methods –
Part 1: General conditions for passenger cars (2019),
Part 2: General conditions for heavy vehicles and buses (2002).
[18] ISO 8855 (2011): Road vehicles – Vehicle dynamics and road-holding ability – Vocabulary.
[19] E.-C. von Glasner: Einbeziehung von Prüfstandsergebnissen in die Simulation des Fahrverhaltens von Nutzfahrzeugen. Postdoctoral thesis, University of Stuttgart, 1987.

Energy demand for drives

Vehicles which draw the energy for propulsion from fossil fuels and consequently emit CO_2 while driving are subject to limits on these emissions. CO_2 emissions are proportional to fuel consumption. This also applies to mild- and full-hybrid vehicles which can run their internal-combustion engines at more fuel-efficient operating points and thus exhibit lower fuel consumption. The batteries of externally rechargeable hybrid vehicles can be charged from the power socket of the mains-supply network or from a charging post; this supplied electrical energy stored in the traction battery must therefore be taken into account in the energy demand. Electric vehicles ultimately draw all the energy required for drive purposes from the mains-supply network and thus do not emit any CO_2 locally.

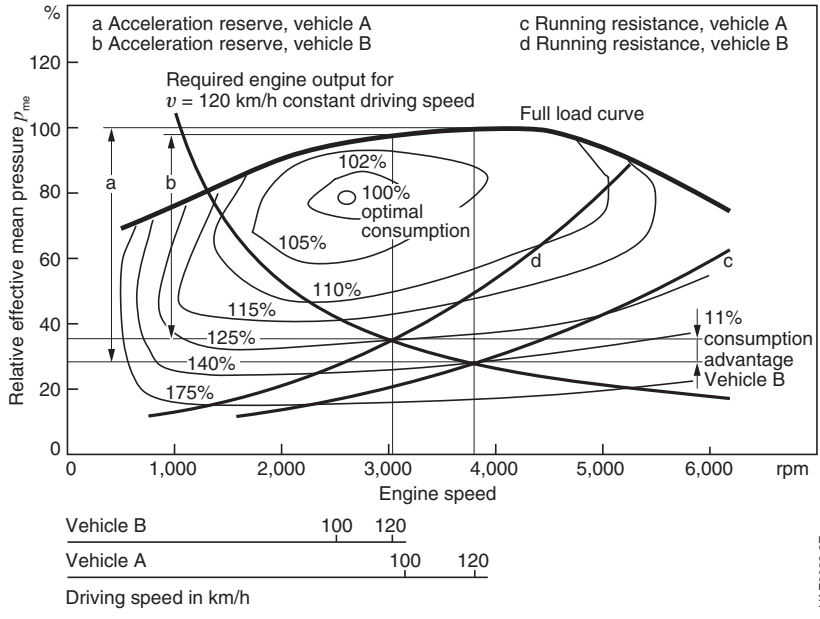
Fuel consumption of internal-combustion engines

Specific fuel consumption

The most common way of presenting engine fuel-consumption maps is the "graph diagram", in which the effective mean pressure p_{me} and, as family parameters, lines of constant specific consumption (fuel throughput referred to power output in g/kWh) are plotted against the engine speed (Figure 1). This makes it possible to compare the efficiency of engines of different sizes and types.

Another way of presentation is, as a family parameter, the fuel throughput or mass flow (e.g., in kg/h). This form of presentation is particularly suitable as an input variable for CAE programs (computer-aided engineering), which can be used to simulate fuel consumption (corresponding to CO_2 emissions).

Figure 1: Engine fuel-consumption map (graph diagram)



Entered in both presentations as additional information are the full-load torque curve as the upper limiting line, the idle and breakaway speeds as limiting speeds, and often also lines of constant power P (power hyperbolas on account of $P \sim p_{me} n$, with the engine speed n).

Consumption as sum of all losses and running resistances

If one disregards the influence of the driver on fuel consumption, which can run to 30 %, there are, based on the consumption formula (Figure 2), three distinct groups of influencing factors:

Table 1: Quantities and units

Quantity		Unit
B_e	Distance consumption	g/m
B_e	Specific fuel consumption	g/kWh
m	Vehicle mass	kg
A	Frontal area of vehicle	m^2
f_R	Coefficient of rolling resistance	–
c_W	Drag coefficient	–
g	Gravitational acceleration	m/s^2
t	Time	s
v	Driving speed	m/s
A	Acceleration	m/s^2
B_r	Braking resistance	N
$\eta_{ü}$	Transmission efficiency of the drivetrain	–
ρ	Density of the air	kg/m^3
α	Gradient angle	°

– the engine (including belt drive and ancillaries),
– the internal running resistances of the drivetrain (e.g., transmission, differentials), and
– the external running resistances.

External running resistances

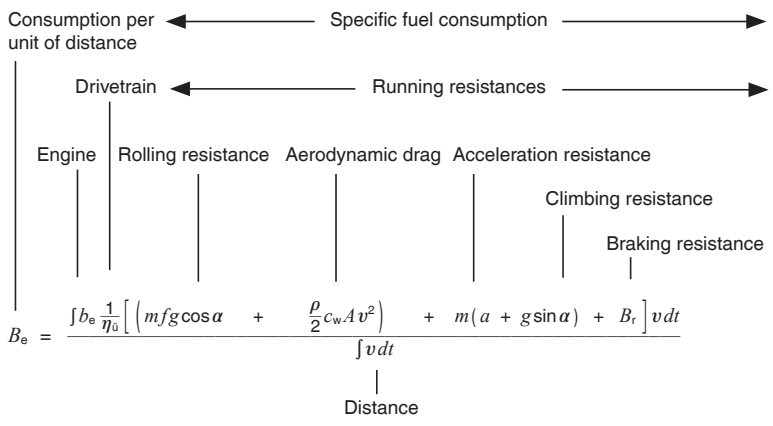
The external running resistances determine a vehicle's minimum energy demand in a specified driving profile. They can be reduced by measures such as reduced vehicle mass, lower tire rolling resistance, and improved aerodynamics. On an average production vehicle without regenerative braking, 10 % reductions in mass, drag and rolling resistance result in fuel-consumption reductions of roughly 6 %, 3 %, and 2 %, respectively.

The formula in Figure 2 distinguishes between acceleration resistance and braking resistance. This illustrates clearly that consumption per unit of distance increases above all when the brakes are subsequently applied and not the overrun fuel cutoff of the engine or even a hybrid system is used to recover some of the brake energy.

Internal running resistances

The internal running resistances are made up of the losses in the drivetrain from the crankshaft to the wheels. In Figure 1, the

Figure 2: Effect of vehicle design on fuel consumption
Quantities, see Table 1.



sum total of the internal and external running resistances are shown by curves c and d, where vehicle A shows significantly lower resistances than vehicle B.

As well as the transmission losses in the drivetrain, the overall ratio also affects fuel consumption. This is calculated from the product of the transmission and differential ratios. The choice of overall ratio results in different operating points in the engine fuel-consumption map for a specific driving speed. A “longer” ratio, i.e., a smaller overall ratio, will generally move the operating points to map areas corresponding to lower fuel consumption. At the same time, it must be noted that the accelerating performance is reduced and the NVH behavior worsens (“Noise Vibration Harshness”, impacting on driving comfort). A useful choice of ratio is therefore only possible within limits.

Determining the distance-specific fuel consumption of conventional drives
The official data for standard fuel consumption are determined on the basis of dynamometer tests conducted on exhaust-emission test benches in which statutory test cycles (e.g., NEDC for Europe, FTP 75 and Highway for the USA and JC08 for Japan) are run. WLTC (World Light-duty Test Cycles), intended to deliver more realistic consumption figures, have been applied since 2017 and 2018 in Europe and Japan. The exhaust gas is collected in sample bags and its constituents HC, CO and CO_2 are subsequently analyzed for the purpose of determining consumption (see Exhaust-gas measuring techniques). The CO_2 content of the exhaust gas is proportional to the fuel consumption.

The following reference values apply to Europe:
Diesel: 1 *l*/100 km ≈ 26.3 g CO_2/km.
Gasoline:
(Euro 4): 1 *l*/100 km ≈ 24.0 g CO_2/km.
(Euro 5, E5): 1 *l*/100 km ≈ 23.4 g CO_2/km.
(Euro 6, E10): 1 *l*/100 km ≈ 22.8 g CO_2/km.

The change from Euro 4 to Euro 5 and Euro 6 is based on a 5 % admixture of ethanol in gasoline (E5) and a 10 % admixture for Euro 6 (E10).

Vehicle masses are simulated on the test bench as an alternative by test masses. In today's certification cycles the test mass is, depending on the country and the curb weight of the vehicle, graded in increments of 55 to 120 kg. When the vehicle mass is allocated to the corresponding the ready-for-operation weight of the vehicle (including all the fillers, tool kit and the fuel tank filled to 90%) plus 100 kg as substitute for a driver and luggage is applied. The difference in consumption in the jump to an adjacent inertia weight class is – depending on the vehicle – between 0.15 and 0.25 *l*/100 km. In the WLTC the actual vehicle mass is applied as the test mass.

Units of fuel consumption
Standard fuel consumption is given in different units, depending on the country and the test cycle. In Europe, the figure is given in g CO_2/km or *l*/100 km, in the USA in mpg (miles per gallon), and in Japan in km/*l*.

Conversion examples:
30 mpg → 235.21/30 → 7.8 *l*/100 km,
22.2 km/*l* → 100/22.2 → 4.5 *l*/100 km.

The fixed value of 235.21 used for conversion is obtained from the conversion of gallons to liters and the conversion of miles to kilometers.

Fuel consumption of hybrid vehicles

Particularities of fuel consumption of autarkic hybrids

In the dynamometer testing of autarkic hybrids (i.e., hybrids without external electric charging) the battery current is determined as well as the CO_2 value and it is ensured in all the country-specific legislation that the battery state of charge is the same before and after the test. The permitted tolerance and possible corrections differ between Europe, USA, and Japan.

In hybrid vehicles some of the vehicle's kinetic energy is converted by recuperation into electrical energy and returned to the battery, where it is available for propulsion to assist the internal-combustion engine (boost mode) or for purely electric driving. The internal-combustion engine's operating point can be moved by the electric machine into ranges in the engine map constituting lower fuel consumption (Figure 1); unfavorable part-load ranges are avoided entirely. In this way, fuel savings of 10 % to 20 % compared with conventional vehicles can be achieved.

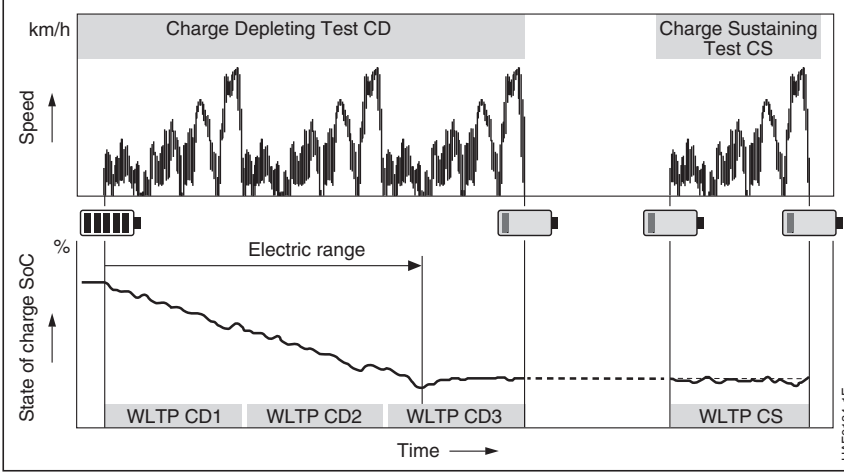
Particularities of fuel consumption of externally rechargeable hybrids

Externally rechargeable hybrids (e.g., plug-in hybrids) differ from autarkic hybrids in that the battery can be charged externally at a power socket or at a wall charging station (wallbox). Wall charging stations enable the battery to be charged at high power levels and thus reduce the charging time.

In contrast to autarkic hybrids and conventional vehicles, two dynamometer tests are performed on externally rechargeable hybrids to determine consumption and the consumption results for fuel and electrical energy from the power socket are offset with different weighting depending on the legislation. Here a test is conducted with a fully charged battery and the cycle is repeated until the electric storage device has reached its minimum charge. This is called the depleting test. A second test, the sustaining test, is performed at the minimum-charge level and is the same as the test for autarkic hybrids (Figure 3).

The utility factor (UF) weights the fuel consumption with a charged storage device (charge-depleting test, CD) and a flat storage device (Charge-sustaining test, CS) as a function of the electric range. It can be determined from field data in accordance with SAE standard J2841 [1]

Figure 3: Test procedures, example WLTC

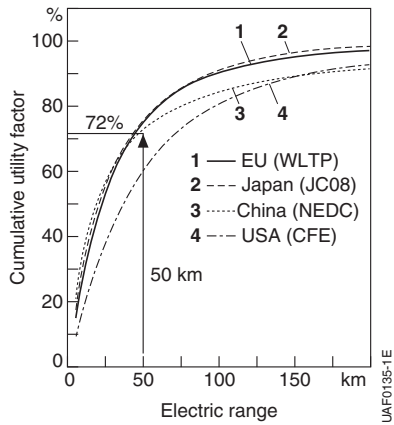


or is defined by the legislature. Figure 4 shows a comparison of the cumulative utility factors for different countries (EU, USA, China, Japan) referred to the WLTP range (Worldwide harmonized Light vehicles Test Procedure). The higher the curve, the more the fuel consumption is weighted in the CD test compared with the CS test consumption for the same range, the utility factor increases, and the certification consumption decreases. This is obtained from the product of sustaining consumption and the value (1– UF).

Figure 4 shows that for an electric range in the WLTC of 50 km the utility factor in China is 72 %, i.e., for sustaining consumption of 6 l/100 km the certification consumption would be 1.7 l/100 km.

The fuel consumption of a plug-in hybrid on the road can deviate markedly from the certification values since the share of electric driving depends, as well as on the driving style, very much on the customer's electric-charging behavior. The charging behavior is decisively influenced by the existing charging infrastructure and the operating convenience.

Figure 4: Utility factor: as a function of the electric range in the WLTC



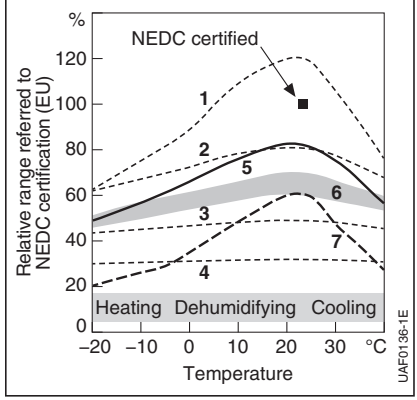
Range and consumption of electric vehicles

Electric range

The official data for the standard range are determined on the basis of dynamometer tests on test benches in which statutory test cycles are repeatedly run. For example, for an electric range of 200 km the NEDC cycle must be run approximately 19 times. As in the dynamometer tests conducted on conventional vehicles and hybrids, secondary loads and air-conditioning influences are not taken into account here either. The latter play a large role in electric vehicles with regard to the attainable range because all the energy for heating and cooling is drawn from the battery and thus is no longer available for propulsion. Figure 5 shows the air-conditioning influence for heating, cooling or dehumidifying (reheating), represented by the temperature, and the influence of driving behavior (speed, acceleration, braking response), represented by different cycles, referred to the attainable range.

Figure 5: Relative range change as a function of temperature referred to the NEDC certification range

1 Constant 50 km/h,
2 Constant 100 km/h,
3 Constant 150 km/h,
4 Constant 200 km/h,
5 WLTC,
6 Long distance,
7 NYCC (urban).



The higher the energy demand from the driving task, the less the effect of the A/C secondary loads. The smallest ranges are attained in urban cycles (using the example of the New York City Cycle, NYCC) at low temperatures. Constant-speed driving at temperatures around 20 °C provide for the largest ranges.

Power-socket consumption

After the end of the range test the traction battery of each electric vehicle is recharged at a power socket or a wall charging station and the energy required for this is determined. Referring this energy to the previously determined range produces the power-socket consumption in Wh/km or kWh/100 km. While the terminal consumption, which can be measured at the battery terminal, is only dependent on the vehicle and the test cycle, the power-socket consumption additionally varies as a result of the charging procedure used. Figure 6 plots the dependence of the power-socket consumption on charging power and charging mode (charging procedure) for a concrete pair of vehicle and driving cycle. The different charging modes are alternating-current charging (AC) with Mode 2 (charging from domestic power socket, with max. 32 A) and Mode 3 (charging e.g. from charging post or wall charging station) and direct-current charging (DC). The higher the charging power, the lower the power-socket consumption and the higher the charging-system efficiency η. This is calculated as follows:

$$\eta = \frac{\text{power-socket consumption in } \frac{\text{kWh}}{100\ \text{km}}}{\text{terminal consumption in } \frac{\text{kWh}}{100\ \text{km}}}$$

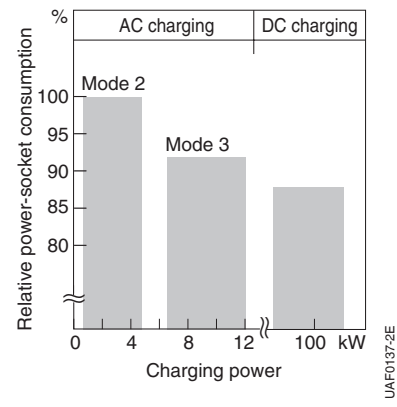
The charging-system efficiency is affected by the losses from ancillary components (e.g., pumps, fans, and ECUs), the losses from the on-board charger during AC charging, and those from the wall charging station. The on-board charger here has the task of converting the alternating current from the charging network into direct current with the right voltage for the battery.

Driving performance of electric vehicles

The driving performance of electric vehicles is much more heavily dependent on temperature and the charge state of the energy accumulator than that of conventional vehicles. As the temperature decreases and with a low battery state of charge (SoC), the electric drivetrain delivers less power and consequently the top speed decreases and the acceleration time increases. At temperatures below freezing and with the battery at 50 % charge the maximum speed decreases to approximately 50 % of the nominal top speed, and the acceleration time from 0 to 100 km/h doubles.

Figure 6: Electrical power-socket consumption as a function of charging power and charging procedure

Relative power-socket consumption is referred to the consumption at 1.4 kW charging power.



Well-to-Wheel analysis

In a Well-to-Wheel analysis the energy consumption and the greenhouse-gas emissions which are caused by the manufacture, provision and use of a fuel are analyzed. The analysis is divided into two steps (Figure 7):

– The Well-to-Tank path describes the provision of fuel.
– The Tank-to-Wheel path describes the use of the fuel in the vehicle.

Provision

The Well-to-Tank path (WtT, provision path) comprises the extraction of the primary sources of energy, the transportation of the raw materials, the manufacture of the fuel, and its distribution onward to the vehicle tank. Both energy losses and greenhouse-gas and exhaust emissions arise in this step.

The Well-to-Tank efficiency (provision efficiency) is the ratio of energy content of the fuel in the tank to the energy content of the source of energy in the deposit. The amounts of energy required for the individual process steps arise as conversion losses (examples, see Table 2).

Different processes must be taken into account here, depending on the fuel. If there are different manufacturing paths for a fuel (e.g., hydrogen), the balance crucially depends on which manufacturing path is chosen. In other words, a fuel cannot be evaluated independently of how it is manufactured. The following examples show specific influencing factors for some fuels.

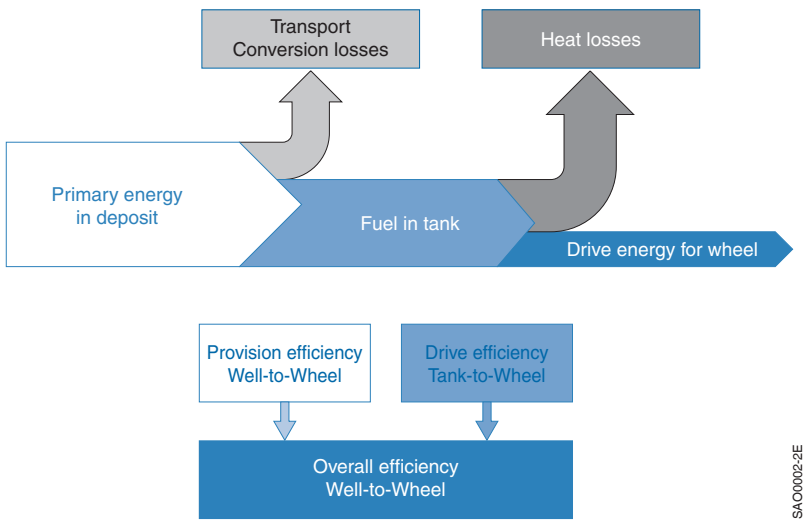
Fossil gasolines and diesel fuels

For the provision of gasoline and diesel fuel the analysis includes above all the energies and emissions which are required for extracting the crude oil, for transporting, processing and refining the oil, and for distributing the fuel.

Biofuels

In the case of biogenous fuels, energy consumption and greenhouse-gas emissions arise during cultivation provided the biomass is cultivated specifically for energy use (cultivated biomass). Particularly crucial here are the manufacture and use of nitrogenated fertilizers as well as diesel consumption for working the cultivated areas. Agricultural production is not taken into account in the biofuels are produced from recyclings.

Figure 7: Energy flow, Well-to-Wheel for an internal-combustion engine (shown as a Sankey diagram)



In the case of fuels produced from biomass, the amount of CO_2 taken by the plants from the atmosphere as they grow are taken into account in the balance as a negative value. In this way the total amount of CO_2 which is attributed to provision is reduced.

Credits
During the production of biofuels by-products are obtained from time to time which can be used for other purposes, e.g., glycerin in the manufacture of biodiesel or animal feed in the manufacture of bioethanol from grain. Provided the by-products replace conventionally created products, saved emissions and energy consumption for the conventional manufacture of the product are credited to the biofuel. In this way energy consumption and emissions which are attributed to the biofuels are reduced. All in all, therefore, negative values can also be obtained for the CO_2 emissions.

In the case of residual biomass, it must be borne in mind that this biomass under certain circumstances could have been used for other purposes (e.g., used frying oil as animal feed, offal timber for generation of energy). This lost alternative use reduces the advantages of the biofuel in the balance.

Natural gas
The provision path of natural gas is essentially described by the extraction, long-distance transportation and compression of the gas to 200 bar at the filling station.

Electricity
If electricity is used directly for the vehicle drive or indirectly to create hydrogen or synthetic fuels, the current-generation emissions are attributed to the fuel.

Hydrogen
Hydrogen can be obtained by steam reformation from natural gas. What must be taken into account in particular here are the long-distance transportation of the natural gas, steam reformation, and compression of the hydrogen at the filling station.

If, however, hydrogen is obtained by electrolysis from water, the electrical energy used must be taken into consideration. The balance is then very much dependent on how the electrical energy is generated, since the emissions from energy generation are attributed to the hydrogen.

Table 2: Conversion losses and provision efficiencies for diesel, gasoline, CNG (compressed natural gas), and H_2 (compressed hydrogen) from two different production paths (gray hydrogen from CNG by means of steam reformation, green hydrogen from wind power by means of electrolysis). Source: [2]

Losses and overall efficiency referred to source energy		Diesel fossil (EU mix)	Gasoline fossil (EU mix)	CNG fossil (EU mix)	H_2 from CNG (Russia)	H_2 Electrolysis (wind power)
Losses	Extraction, energy generation, processing	10 %	10 %	2%	2%	0 %
	Long-distance transportation	1 %	1 %	4 %	8 %	0 %
	Conversion (refinery, steam reformation, electrolysis)	8 %	7 %	0 %	26 %	32 %
	Domestic distribution	1 %	2%	7 %	10 %	15%
Overall provision efficiency		79 %	81 %	87 %	54 %	54 %

Synthetic fuels
The synthesis of hydrogen into gaseous or liquid fuels can essentially be performed in one of two processes: methanation or by way of a Fischer-Tropsch process. If for these processes CO_2 is filtered out of the air and the hydrogen is obtained from regenerated electricity, the fuel produced can in the best case be greenhouse-gas-free.

The generation of the starting substances hydrogen and CO_2 or CO and the number of subsequent synthesis steps determine the energy losses which arise during the production of the various synthetic fuels (e-fuels).

Drive
The fuel used and the drive efficiency are decisive for the fuel consumption and the local CO_2 emissions of a vehicle. This is referred to as the Tank-to-Wheel efficiency and describes the ratio of drive energy which is output to the wheels (Tank-to-Wheel path, drive path) to the energy content of the fuel consumed.

Defined driving cycles are used to compare the fuel consumption and the emissions of different drive concepts. This is in effect a driving profile with defined driving speeds and acceleration phases. For a standardized driving cycle the respective drive efficiency and the emissions can be determined for the different drive concepts.

Overall examination
The provision path and the drive path are brought together in the Well-to-Wheel analysis (WtW) (see Figure 7). Here in each case the energy expenditure and the greenhouse-gas emissions of the two sub-paths are added. The overall examination thus comprises both fuel provision and vehicle use. In the energy balance the sum total of exhaustible primary energies is given in MJ/km (i.e., fossil and nuclear energy). In the greenhouse-gas balance the sum total of greenhouse-gas emissions is given as a CO_2 equivalent in g/km.

The CO_2 equivalent takes into account as well as carbon dioxide further greenhouse gases, which are weighted according to their specific greenhouse effect. Reference vehicles and driving cycles are used as the basis for the balances.

When the different fuels are compared with each other, it becomes clear that a relatively high expenditure for fuel provision (energy used and emissions caused) can be compensated for by a high drive efficiency and low direct emissions during the combustion of the fuel. It is therefore not possible to evaluate a fuel without also analyzing the associated drive technology.

Figure 8 shows by way of example for different fossil and regenerative fuels the CO_2 emissions which result from fuel provision (WtT) and from fuel use (TtW). It must be borne in mind here, however, that the CO_2 balance is very much dependent on the boundary conditions of fuel production and use such that the specified values are in no way generally applicable to the fuel in question.

Life-cycle examination
A further analysis of greenhouse-gas emissions and energy consumption moves the focus of examination from the fuel to the vehicle. A life-cycle examination for the vehicles takes into consideration, aside from operation of the vehicle, energy consumption and greenhouse-gas emissions that arise during the manufacture of the vehicle. Starting with raw-material extraction, transportation, refinement, and processing, here too recycling is taken into consideration at the end of the vehicle's life (“cradle to grave”). The results of this examination are very much dependent on the process paths used.

A more comprehensive examination of the environmental impacts is offered by a life-cycle analysis (LCA). Such an analysis takes into account, as well as energy consumption and greenhouse-gas emissions, numerous additional environmental impacts such as for example land consumption, water consumption, and water pollution.

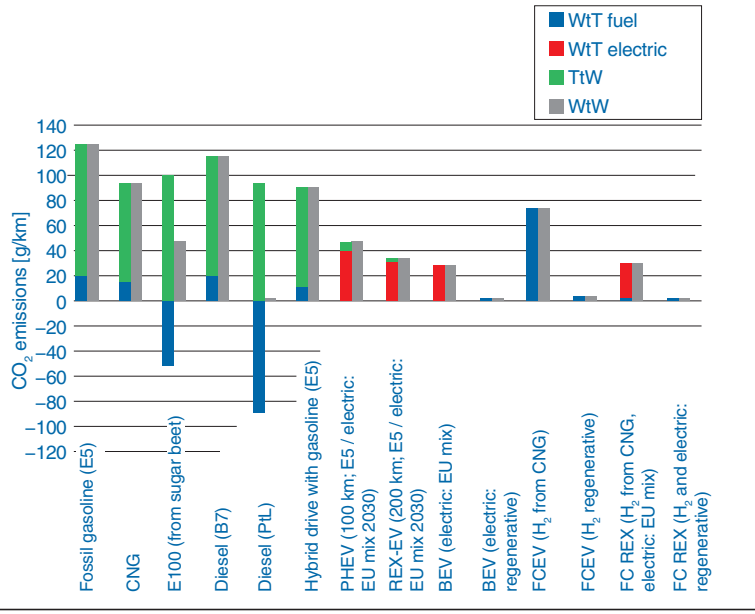
References

[1] SAE J2841: Utility Factor Definitions for Plug-In Hybrid Electric Vehicles Using Travel Survey Data.

[2] JEC Technical Report, 2020; Appendix 1 Version 5.

Figure 8: Greenhouse-gas emissions from the provision and utilization of different fuels in a vehicle with internal-combustion engine or alternative drives. Source: [2]

WtT Well to Tank,
TtW Tank to Wheel,
WtW Well to Wheel,
E100 Ethanol, E5 Gasoline with 5 % ethanol additive, B7 Diesel fuel with 7 % biodiesel additive,
PtL Power to Liquid (liquid synthetic fuels),
CNG Compressed Natural Gas,
PHEV Plug-in Hybrid,
REX-EV Battery Electric Vehicle with Range Extender,
BEV Battery Electric Vehicle,
FC Fuel Cell,
FCEV Fuel Cell Electric Vehicle, H_2 is produced by electrolysis, electrical energy is produced in accordance with present-day EU mix or regeneratively,
FC Range Extender: FCEV whose battery can be electrically recharged.



Vehicle aerodynamics

Aerodynamic forces

Vehicle aerodynamics concerns all phenomena that arise as a surrounding medium (air) flows around or through moving vehicles and their components. The drag, forces and moments (see Table 1) are evaluated and usually optimized by changing the shape of the vehicle.

Aerodynamic drag
Aerodynamic drag W_L is given a lot of attention in developing motor vehicles, because it has a direct impact on the vehicle performance and fuel consumption. A critical variable here is the drag coefficient c_W, which describes the aerodynamic quality of the shape of a body in a flow. The aerodynamic drag is also influenced by the frontal area A_{fx} and the dynamic pressure q of the flow, which depends on the air density ρ and the airflow velocity v.

Dynamic pressure is described by this equation:

$$q = 0.5\ \rho\ v^2.$$

Aerodynamic drag is calculated as (also refer to fluid mechanics):

$$W_L = \frac{\rho}{2}\, v^2 c_W A_{fx}.$$

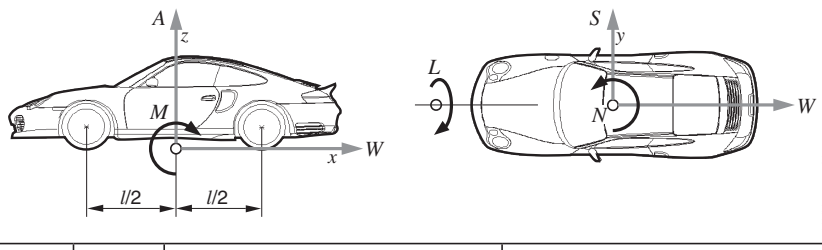
Aerodynamic drag is composed of the skin friction drag and form drag:

$$W_L = W_R + W_D.$$

Skin friction drag W_R
Skin friction drag arises due to shear stress τ_w on the walls as the air flows by. Figure 1 shows an example of this on a wing. Skin friction drag is calculated as an integral of the shear stress over the surface:

$$W_R = \int \tau_w \cos\varphi\, dF.$$

Table 1: Aerodynamic forces and moments
q Dynamic pressure, ρ Air density, v Airflow velocity, A_{fx} Frontal area,
c_A Lift coefficient, l Wheelbase.
$c_{AV} = 0.5\ c_A + c_M$ (relative to the front axle),
$c_{AH} = 0.5\ c_A - c_M$ (referred to rear axle).



Quantity	Unit	Meaning	Coefficient with definition
W	N	Drag in the x direction	$c_W = W/(q\,A_{fx})$
S	N	Side force in the y direction	$c_S = S/(q\,A_{fx})$
A	N	Lift in the z direction	$c_A = A/(q\,A_{fx})$
L	Nm	Rolling moment around x	$c_L = L/(q\,A_{fx}\,l)$
M	Nm	Pitching moment around y	$c_M = M/(q\,A_{fx}\,l)$
N	Nm	Yawing moment around z	$c_N = N/(q\,A_{fx}\,l)$

Form drag W_D
Form drag arises through pressure changes on blunt bodies caused by shedding. It is calculated as an integral of the pressure over the surface:

$$W_D = \int p \sin\varphi \, dF.$$

Figure 2 shows the flow pattern for flow around a wing and a hemisphere. Shedding results in vortices, and therefore pressure differences. Form drag is predominant for blunt bodies. For streamlined bodies, the friction component dominates; there is no turbulent shedding from the body's surface.

For the c_W values for different bodies and car body shapes, refer to the various tables (see Drag coefficients, Fluid mechanics and Drag coefficients, Motor-vehicle dynamics).

Figure 1: Formation of skin friction drag from airflow
v_∞ Airflow velocity,
φ Angle of impact to the surface element,
τ_W Shear stress,
p Pressure,
dF Surface element.

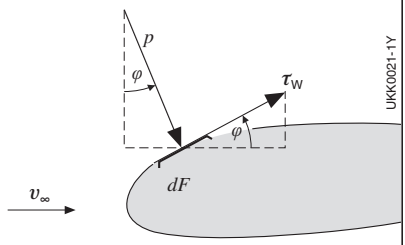
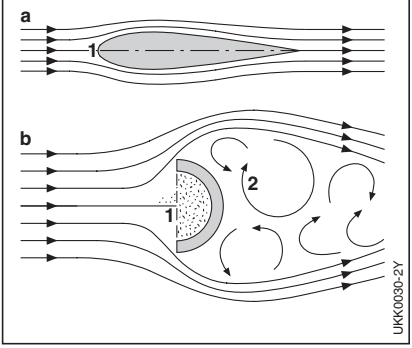


Figure 2: Flow conditions for various bodies
a) Streamlined body (wing),
b) Blunt body (hemisphere).
1 Stagnation point,
2 Separated flow, vortices.



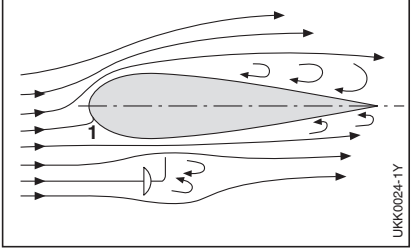
Induced drag
The difference in pressure above and below a vehicle in flow lead to superimposition of the main horizontal flow around the body. A vertical flow component arises through pressure equalization via the body sides, and wake turbulence develops, which causes the induced drag.

Internal drag
Internal drag arises from pressure loss as air flows through the engine compartment.

Interference drag
Interference drag is produced by the interaction of attachments (such as the wheel suspension, wheels, exterior mirrors, antennas, windshield wipers, spoilers, vanes). The total drag is not the sum of the drag forces of the individual parts, because the mutual interference has to be taken into account. Thus attachments can also cause the c_W value to drop.

The impact of an exterior rearview mirror on the car body is shown here as an example (Fig. 3). It causes the stagnation point to move toward the attachment; this results in asymmetrical flow around the car body. Consequently, there is a greater risk of shedding vortices, and that leads to an increasing c_W value.

Figure 3: Influence of an exterior mirror on a car body (model)
1 Stagnation point.



Buoyancy force
The curvature of the car's top surface causes the air to flow faster above the vehicle than under it, which generates an undesirable lift force (Fig. 4). This reduces the tire contact forces and thereby has a negative impact on the directional stability.

The lift coefficient c_A is the sum of the lift coefficients on the front axle c_{AV} and the rear axle c_{AH}. The difference between the lift on the front and rear axles is referred to as "lift balance" and is a variable influencing directional stability.

The pitching moment M, which acts around the y-axis, is often used as a design variable instead. A positive pitching moment requires understeer, and a negative pitching moment requires oversteer. The rolling moment L around the x-axis is disregarded with respect to aerodynamics.

Lateral force
A car has an almost symmetrical shape when viewed from the front, which means that side forces generated by air flows are small. As soon as the approaching air flow deviates from the x-axis (i.e. in a crosswind), the airflow generates side forces that can affect vehicle behavior considerably (Fig. 5).

The yawing moment N, which acts around the z-axis, is also used as an indicator for crosswind sensitivity. This value is used to derive the yaw-angle velocity and yaw angular acceleration, which provide information about the negative impact of crosswind.

Figure 4: Lift resulting from the airflow
v_o Velocity of the airflow over the vehicle,
v_u Velocity of the airflow between the vehicle and road,
p_o Pressure above the vehicle,
p_u Pressure below the vehicle,
A Lift.

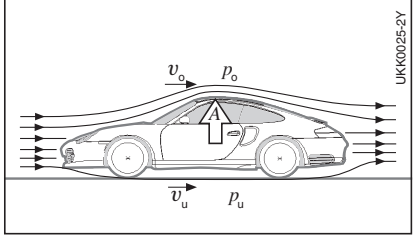
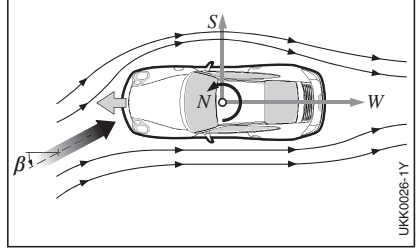


Figure 5: Side forces
S Side force,
W Drag in the longitudinal direction,
N Yawing moment,
β Angle of impact.



Tasks of vehicle aerodynamics

Aerodynamics has other tasks besides reducing the c_W value (Fig. 6).

Motive power

Impact of velocity on tractive resistance

In order to set the vehicle in motion and keep it moving, the engine has to produce a tractive force Z. This is composed of the aerodynamic drag W_L, rolling resistance W_R, climbing resistance W_S, drag during acceleration W_B and losses in the powertrain W_A. This results in the tractive resistance equation:

$$Z = W_L + W_R + W_S + W_B + W_A.$$

Climbing resistance and acceleration resistance are zero during constant motion on a level stretch. The mileage is substantially determined by the aerodynamic drag, because it increases quadratically with the velocity and surpasses the other tractive resistances as of 80 km/h (Fig. 7).

The lower the product of the c_W value and frontal area A_{fx} ($A_{fx} \cdot c_W$ is often called the drag area), the faster a vehicle can move with a suitable transmission ratio. At a constant velocity, a smaller drag area leads to lower fuel consumption. Consequently, the vehicle produces fewer emissions.

As a function of the low driving speeds in the New European Driving Cycle (average speed of 33.4 km/h, see European test cycle), however, the influence of the aerodynamic drag is greatly underestimated in the manufacturer/consumption data, which can lead to higher fuel consumption in everyday operation for the average driver.

Figure 7: Tractive resistance

W_L Aerodynamic drag,
W_R Rolling resistance,
W_A Losses in the powertrain.

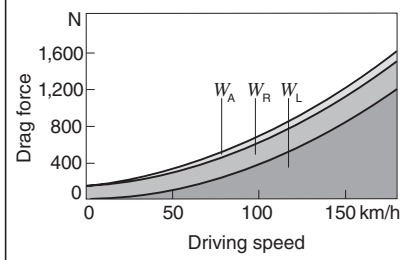


Figure 6: Vehicle aerodynamics: requirements and influences

Vehicle performance	Design	Component forces
Consumption, emissions, top speed, acceleration, driving dynamics.	Design that is characteristic of the brand, visible or invisible aerodynamic measures (retractable rear spoiler).	Doors and cover, windows and sliding sunroof, vibration of mirrors, rear window fluttering, puffing up behavior of convertible top.
Convenience	**Cooling and ventilation**	**Directional stability**
Draft prevention (mobile roofs), wind noise, prevention of dirt buildup.	Ventilation of brakes, engine compartment and components, engine cooling, cooling of units (such as gearbox), intercooling, air conditioning, removing condensation from headlights.	Lift forces, lift balance, straight-running stability, lane-change behavior, crosswind stability, handling.

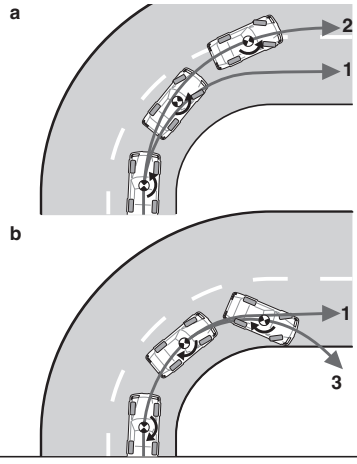
UKK0027-3E

Directional stability
Directional stability, sensitivity of the steering behavior, lane-change behavior, handling and crosswind sensitivity are critically influenced by the lift and side forces. The forces that can be transferred from the wheels to the road in the longitudinal and transverse directions can be no larger than the vertical force (normal force).

The lift acts against the vertical forces resulting from the weight. Consequently, the forces that can be transferred from the tires, above all during motion in a curve, can be reduced until there is a total loss of directional stability, depending on the driving speed and lift. The transferable braking forces also depend on the vertical wheel load, and therefore on the aerodynamic lift.

The task of aerodynamics is therefore to minimize the lift forces through means such as spoilers on the front and rear of a vehicle. For sports cars it can even happen that downforces (i.e. negative lift forces) greater than the weight are produced.

Figure 8: Cornering behavior
a) Understeering behavior,
b) Oversteering behavior.
1 Correct trajectory,
2 Trajectory of an understeered vehicle,
3 Trajectory of an oversteered vehicle.



Lift balance
If the lift balance has a lift coefficient c_{AV} on the front axle greater than c_{AH} on the rear axle, there is a positive pitching moment and the design produces understeering driving behavior. Then if the speed in a curve is too high it will be noticed immediately when the steering angle is too high; the vehicle should be stabilized again by backing off the accelerator (Fig. 8a).

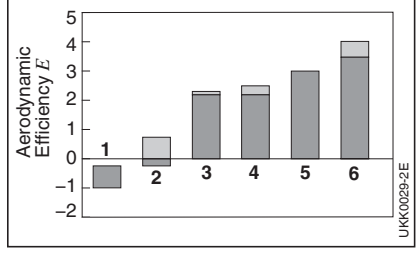
If c_{AV} is less than c_{AH}, a negative pitching moment is produced and there is an oversteering design (Fig. 8b). Then if the speed in a curve is too high it will be perceived not through the steering wheel, but through the seat. Backing off the accelerator makes the vehicle unstable; it can be kept in the lane only with targeted countersteering and depressing the accelerator.

Aerodynamic efficiency
A significant reduction of the lift forces usually comes at the expense of a favorable c_W value. The smaller this impact, the greater the aerodynamic efficiency E. It describes the relationship between the lift and drag coefficients:

$$E = \frac{c_A}{c_W}.$$

Figure 9 shows a comparison of the values for the aerodynamic efficiency of different vehicles. Due to the freedom of body design and the very fast racetrack, current Le Mans prototypes have the best

Figure 9: Aerodynamic efficiency for various vehicle types
1 Passenger car,
2 Sports car,
3 Formula 3 racecar,
4 Indy car series,
5 Formula 1 racecar,
6 Le Mans racecar.



efficiency (E = 3.5 to 4). Despite a high downforce, a low c_W value is achieved.

Due to the fuel consumption advantages, passenger cars usually have very low c_W values, but positive lift, and therefore a negative efficiency down to $E = -1$.

Cooling and ventilation

The responsibility of aerodynamics consists in requirements-based air intake, so that the heat generated by operating the vehicle is dissipated to the environment (see Engine cooling).

An intensive flow of cooling air can lead to a high internal drag, and therefore increase the aerodynamic drag. Therefore a great deal of attention is given to achieving an optimum balance between cooling requirements and aerodynamics (Fig. 10). To optimize the air intake, air inlets and outlets are located on the vehicle in zones with the greatest possible pressure difference for propulsion. The flow of cooling air can be controlled and adapted by changing the size of the respective openings. The cooling air is conducted through channels to the radiators, which should be designed to be as free of pressure loss and leaks as possible.

The same criteria also apply to an exhaust air duct up to the air outlet. The air outlet should be positioned downstream with the least possible pressure loss. If it can be placed in a zone where there are already aerodynamic losses, these can be favorably reduced by the arising interference. For example, blowing out the cooling air ahead of the front wheel can form an air cushion and thereby guide the flow around the wheel with fewer losses, which nearly equalizes the cooling air drag (Fig. 11).

The aerodynamic pressure losses can also be controlled by choosing suitable radiator cores and their dimensions, as well as fans and their frames. To entirely prevent cooling air drag with a low power requirement, electrically closable radiator shutters were introduced in series production back in 1987 (for the Porsche 928). Today this measure is widely used in the premium class.

Figure 11: Interference from blowing out the cooling air ahead of the front wheel

1 Airflow approaching the radiator,
2 Airflow approaching the wheel,
3 Radiator,
4 Exhaust flow after the radiator and air cushion ahead of the wheel.

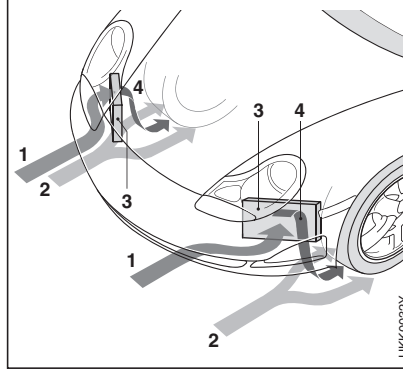
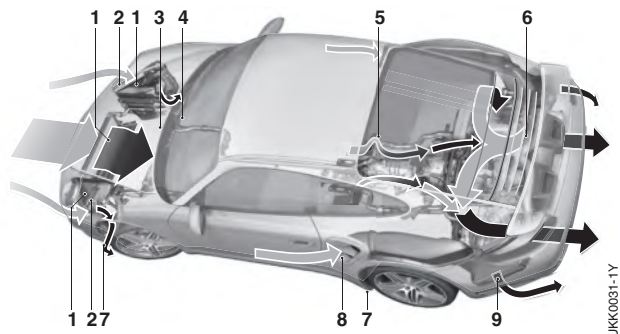


Figure 10: Cooling tasks for a sports car

1 Flow through radiator (engine cooling),
2 Air conditioning,
3 Front axle gearbox cooling,
4 Fresh air intake,
5 Rear axle gearbox cooling,
6 Engine compartment ventilation,
7 Brake ventilation,
8 Intercooling,
9 Cooling air outlet from the intercooler.



Comfort and convenience
Aerodynamics helps provide the passengers with comfort primarily by reducing wind noise (aeroacoustics), dirt and drafts.

Aeroacoustics
Wind noise in the passenger compartment of a car is the primary noise source for speeds over 120 km/h; above 200 km/h it is approximately 6 dB(A) louder than all other acoustic sources (Fig. 12). This corresponds to four times the sound power.

The term "aeroacoustics" refers to the production of noise from
– Flow around the body (flow separation, broadband noise from unsteady pressure fluctuations, noise with discrete frequencies such as from antennas and exterior mirrors),
– Component excitation with low-frequency noise excitation due to large surfaces in a flow (such as convertible top, roof, door and cover panels),
– Overflow with noise created by a Helmholtz resonance (such as with open windows, sliding sunroof),
– Leakage due to pressure differences between the passenger compartment and flow around the body as the result of aerodynamic and mechanical component loads.

Improvement of the aeroacoustics
By giving the basic body an aerodynamic shape you can reduce the local flow velocities, influence the flow direction, and prevent or limit turbulent shedding. Generally speaking, whatever reduces the c_W value also leads to lower wind noise.

Freestanding attachments (such as mirrors, wipers, antennas, roof transport system) disrupt the flow around the basic body by creating additional flow separation. Therefore they should be positioned where they will produce the least possible disruption and their shape should be optimized. With component excitation, aerodynamics analyses the necessity and placement of reinforcement plates or damping measures.

In the case of flow over cavities (such as an open sliding sunroof), low-frequency pulsing of the air volume in the passenger compartment can occur (humming). That is the resonance case of periodic oscillation of the shear layer between the air overflow and the static air in the passenger compartment (Fig. 13) with the natural frequency of the passenger compartment (Helmholtz resonator principle). As an aerodynamic measure, the shear layer buildup is upset by interference flows, for example, using a wind deflector on the sliding sunroof.

Figure 12: Wind noise components in total noise
Measurement of the sound pressure level with an artificial head in the driver's seat, left ear.
1 Total noise,
2 Wind noise,
3 Other noise.

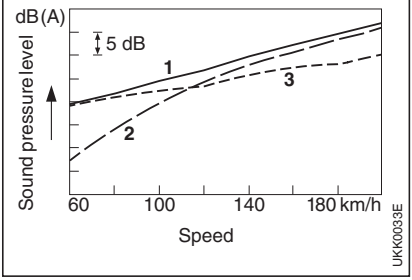
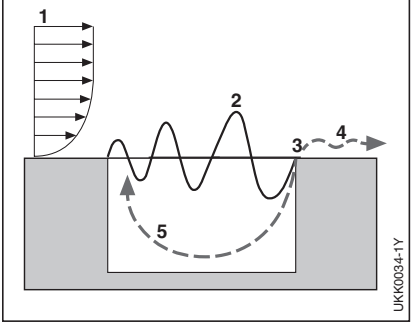


Figure 13: Feedback mechanism
1 Turbulent oncoming flow,
2 Shear layer oscillation,
3 Rear edge of cavity,
4 Vortex,
5 Pressure wave.



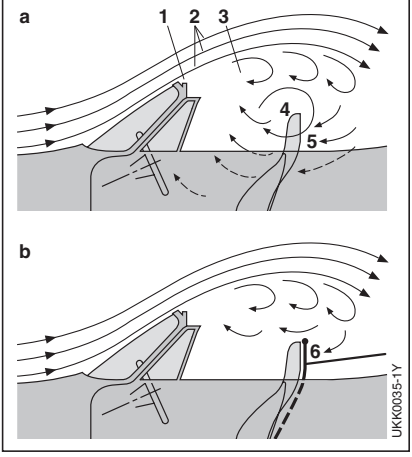
The aeroacoustic quality is evaluated not only by the spectral analysis, but also by taking into account psychoacoustic parameters. These parameters measure higher-frequency noise components; consequently, subjective perception (loudness, sharpness) and speech intelligibility are taken into account better. The sharpness is evaluated independently from loudness and roughness. The articulation index is a measure of speech intelligibility.

Dirt

The wheels of your own vehicle kick up dirt particles, which settle on the surface of your vehicle (self-caused dirtying) or they mix with the turbulent wake and settle on following vehicles (extraneously caused dirtying). The magnitude and intensity of the contaminated surfaces are caused by local wake turbulence of the surrounding flow. The task of vehicle aerodynamics is to identify such turbulence and reduce it or transpose it to unproblematic zones. Great care is given to avoiding impaired visibility by keeping the windows and exterior mirrors free of dirt buildup.

Figure 14: Reduction of drafts for a convertible by using a windbreak

a) Vehicle without windbreak,
b) Vehicle with windbreak.
1 Flow separation at the top edge of the windshield,
2 Flow around vehicle,
3 Unsteady vortex sheet,
4 Vortex direction of rotation in passenger compartment,
5 Airflow between the seats and between the seats and side panels,
6 Windbreak for reducing airflow between the seats.



Draft prevention

Usually there are bothersome drafts when driving with the windows and roof open. This can be clearly observed in a convertible when the top is open, because there is shedding of the flow over a large area behind the windshield frame. A low-pressure region behind the windshield is filled by a backflow. Backflows and turbulence cause the bothersome drafts (Fig. 14).

The vortices can be damped and significantly reduced by the mesh fabric of a windbreak. The task of vehicle aerodynamics is to determine the size and shape of this component as well as its mesh density.

Component forces

Component forces arise from the pressure forces produced by the flow around the body (measured quantity is the c_p value), which vehicle aerodynamics has to determine empirically and increasingly also through calculations. The positive or negative pressure forces acting perpendicular to the outer skin increase along with the airflow velocity and act as a load on the components located there.

In areas where there is shedding (such as A-pillars, exterior mirrors, rear window, rear spoiler), the turbulence results in stochastically varying loads, which have to be taken into account for determining the function and durability of the components.

Vehicle wind tunnels

Application

Aerodynamic measurements on roads are difficult, because the environmental conditions are neither uniform nor steady. The direction and steadiness of the conditions are constantly changing due to factors such as natural wind, buildings, and traffic. A vehicle wind tunnel is used to represent the air flow affecting a vehicle traveling on the road in an environment that is as realistic and reproducible as possible.

The advantages of using a vehicle wind tunnel as an experimental development tool compared with road measurements are the reproducibility of test conditions, the uncomplicated, reliable, and fast measuring technology, and the ability to disassociate effects (such as driving noise) that occur only in conjunction with road travel. Design prototypes that are not yet roadworthy can be optimized aerodynamically in a wind tunnel with guaranteed secrecy.

Wind tunnel designs

Types of wind tunnel

Aerodynamic parameters are calculated in wind tunnels. They vary in the type of air guidance, test section, size, and road-surface simulation (Table 2). Closed-circuit (closed return) wind tunnels are referred to as "Göttingen type", while open-circuit (no return) wind tunnels are referred to as "Eiffel type" (Figure 15). For vehicle aerodynamics, Göttingen wind tunnels with an open test section or with slotted walls are used predominantly.

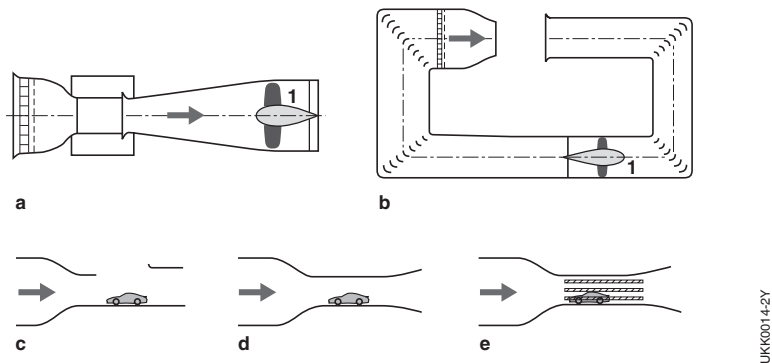
Standard wind tunnel equipment

The test section design can be open, have slotted walls, or be closed. It is characterized by the outlet cross-section of the nozzle, the cross-section of the collector, and the length of the test section (Fig. 19, Table 2).

Obstruction $\Phi_N = A_{fx}/A_N$ is also an important parameter. This is the ratio between the frontal area of the vehicle A_{fx} and the nozzle cross-section A_N. On the road, this ratio is $\Phi_N = 0$, therefore it should also be as small as possible in the wind tunnel. Taking the construction and operating costs of a wind tunnel into account, $\Phi_N = 0.1$ is a common figure in practice. This corresponds to a diffuser cross-section of approx. 20 m^2.

The contraction and contours of a diffuser determine the velocity and steadiness of the wind-tunnel air flow. A large contraction ratio κ of the cross-sectional areas of the prechamber to the nozzle outlet ($\kappa = A_V/A_D$) leads to uniform velocity distribution and low turbulence, as well as to fast acceleration of the airflow.

Figure 15: Wind-tunnel designs
a) Eiffel type,
b) Göttingen type,
c) Open test section,
d) Closed test section,
e) Test section with slotted walls.
1 Fan.



The diffuser contour can influence the steadiness of the velocity profile at the diffuser outlet in the test section, and parallelism in relation to the geometrical tunnel axis.

The settling chamber or prechamber is positioned upstream of the nozzle and in the largest cross-sectional area of the wind tunnel. The prechamber contains flow rectifiers, filters, and heat exchangers that improve the quality of the air flow in terms of steadiness and direction, and keep the channel temperature constant.

The fan in most wind tunnels can generate wind speeds of well over 200 km/h. Speeds of this size are only rarely used, for example, for testing the functional safety and stability of body components under wind loads. This is because such tests require the full fan capacity of up to 5,000 kW. Normal measurements are taken at 140 km/h. At this speed, the aerodynamic coefficients can be determined reliably and at low operating cost. The wind speed is regulated either by changing the fan speed, or by adjusting the fan blades at a constant speed (Table 3).

A wind tunnel balance records the aerodynamic forces at work on the test vehicle and the moments of all four tire contact surfaces; it breaks these forces into x, y, and z components to calculate the aerodynamic parameters.

A wind tunnel balance is normally positioned below a turning platform, which is used to turn the vehicle in relation to the wind direction, and thus simulate the effect of side winds.

In contrast to a real situation, a vehicle in a wind tunnel is stationary and has air blown toward it. The influence of the relative movement between the vehicle and

Table 2: Vehicle wind tunnels in Germany (examples)

Wind tunnel operator	Nozzle cross-section	Collector cross-section	Test section length	Test section design	Simulation of the moving road
Audi	11.0 m²	37.4 m²	9.5 to 9.93 m	Open	with 5 moving belts
BMW	18.0 to 25.0 m²	41.4 m²	17.9 m	Open	with 5 moving belts
BMW Aerolab	14 m²	25.9 m²	15.7 m	Open	1 conveyor belt, width greater than vehicle width
Daimler	28.0 m²	61 m²	19 m	Open	with 5 moving belts
IVK Stuttgart	22.5 m²	26.5 m²	9.9 m	Open	with 5 moving belts
Ford	20.0 m²	28.2 m²	9.7 m	Open	–
Porsche	22.3 m²	37.7 m²	13.5 m	Slotted walls, open	–
Volkswagen	37.5 m²	44.8 m²	10.0 m	Open	–

Table 3: Wind-tunnel fans

Wind tunnel	Drive power	Fan diameter	Wind velocity	Control
Audi	2,600 kW	5.0 m	300 km/h	Revolutions per minute
BMW	4,140 kW	8.0 m	300 km/h	Revolutions per minute
BMW Aerolab	3,440 kW	6.3 m	300 km/h	Revolutions per minute
Daimler	5,200 kW	9.0 m	265 km/h	Revolutions per minute
IVK Stuttgart	3,300 kW	7.1 m	260 km/h	Revolutions per minute
Ford	1,950 kW	6.3 m	185 km/h	Revolutions per minute
Porsche	2,600 kW	7.4 m	220 km/h	Revolutions per minute
Volkswagen	2,600 kW	9.0 m	180 km/h	Blade adjustment

the road cannot be taken into account. Therefore, in addition to wind tunnels with a stationary floor, there is now increased use of wind tunnels with moving belts incorporated into the floor to imitate motion of the car over the road and rotating wheels (Figure 16). The quality of the air flow between the vehicle and the road can thus be improved and made significantly more realistic.

Wind-tunnel auxiliary systems

Frontal-area measuring system

The frontal-area measuring system (laser or CCD system) measures the frontal area of the vehicle by optical means. The results are then used to calculate the aerodynamic coefficients from the forces measured in the wind tunnel.

The ideal pressure transducer system of a wind tunnel can simultaneously record the pressure curve at up to 100 pressure measuring points, for example, using the flat sensors affixed to the car body surface for the pressure distribution. The miniature pressure sensors used for each measuring point have no wearing parts

Figure 16: Turning platform on test section floor with flat belts incorporated

1 Moving belt between the wheels,
2 Balance,
3 Turning platform,
4 Wheel drive unit with small moving belt.

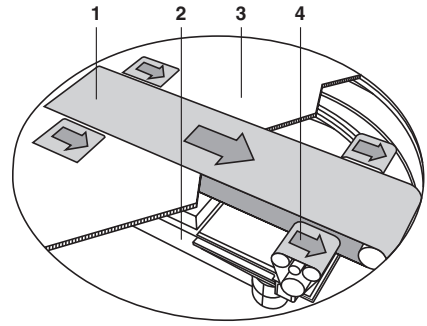
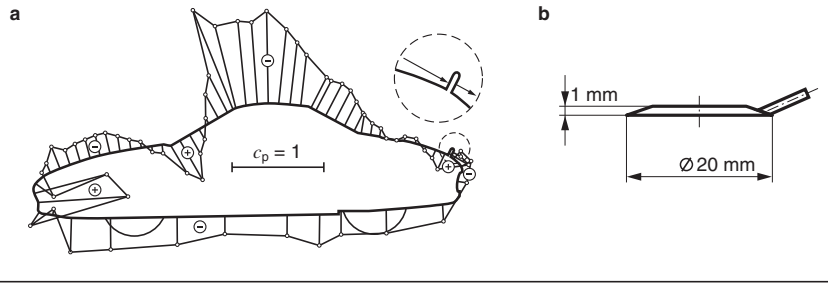


Figure 17: Measurement of pressure distribution (application example)

a) Pressure distribution on midsection of vehicle surface (y = 0), measured using 63 flat sensors,
b) Flat sensor.
The lines show the static pressure acting perpendicular to the measurement point.
The arrow length indicates the magnitude of the pressure.
The line under c_p = 1 shows the scale.
c_p is a dimensionless pressure coefficient;
it describes the pressure p_∞ at any point on the vehicle at the velocity v_∞ with respect to the dynamic pressure p prevalent at this velocity.

$$c_p = \frac{p - p_\infty}{0.5 \cdot \rho \cdot v_\infty^2}.$$



and can therefore electronically scan at a high frequency (Figure 17).

Traversing cradle
The traversing cradle allows the measurement of the complete vehicle airflow field. Each position in the test section can be located by coordinates and reproduced. According to the sensors present at each point, the pressure, velocity, or noise sources at each location can then be determined.

Smoke tubes
Smoke tubes are used to visualize the airflow, which would otherwise be invisible. The smoke pattern enables detection of flow separation, the vortices of which worsen the coefficients (Fig. 18). The nontoxic "smoke" is usually created by heating a glycol mixture in an oil-vapor generator. Other methods for visualizing the air flow include:
– Tufts on the vehicle surface,
– Tuft sensors,
– Photographs of flow patterns with quick-drying liquid mixture of kerosene or talc,
– Helium bubble generator,
– Laser sheet.

Pollutant dispersion system
The pollutant dispersion system can be used to spray the vehicle in the wind tunnel with water ranging from light mist to heavy rain. The flow patterns can be visualized and documented by mixing chalk or a fluorescent agent to the pollutant.

Hot-water unit
The hot-water unit provides heated water at a constant rate for measuring radiator cooling capacity in prototypes that are not yet roadworthy.

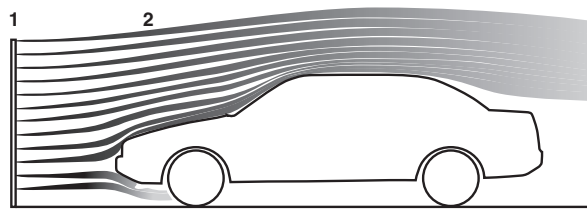
Wind tunnel variants
A vehicle wind tunnel is a major investment. This investment, combined with high operating costs, make it an expensive test environment with a high hourly rate. Only very frequent utilization of vehicle wind tunnels for aerodynamic, aeroacoustic, and thermal experiments can justify the construction of several wind tunnels specialized for different tasks.

Model wind tunnel
A model wind tunnel can considerably reduce operating costs, due to the lower construction requirements and technical complexity. Depending on the scale (1:5 to 1:2), it is easier, faster, and more cost-effective to change the shape of vehicle models.

Model tests are primarily used in the early development phase for optimizing the basic aerodynamic shape. With the support of designers, plasticine models are used in a wind tunnel to optimize shapes or develop complete alternative shape variants and establish their aerodynamic potential.

Once there is a data set, new production methods (such as rapid prototyping) can be used to quickly produce models to any scale with accurate details. This allows for important investigations of the models in order to optimize details, even after the shaping phase.

Figure 18: Use of a smoke rake consisting of smoke tubes to make the airflow visible
1 Rake,
2 Smoke stream.



Acoustic wind tunnel
In an acoustic wind tunnel, extensive sound insulation measures mean that the sound pressure level is approximately 30 dB(A) lower than in a standard wind tunnel. This provides a sufficiently large signal-to-noise ratio of more than 10 dB(A) from the useful signal of the vehicle to allow identification and evaluation of wind noises generated by the air circulation and throughflow.

Climate-controlled wind tunnels
Climate-controlled wind tunnels are used for thermal analysis and protection of vehicles in defined temperature ranges at different load conditions.

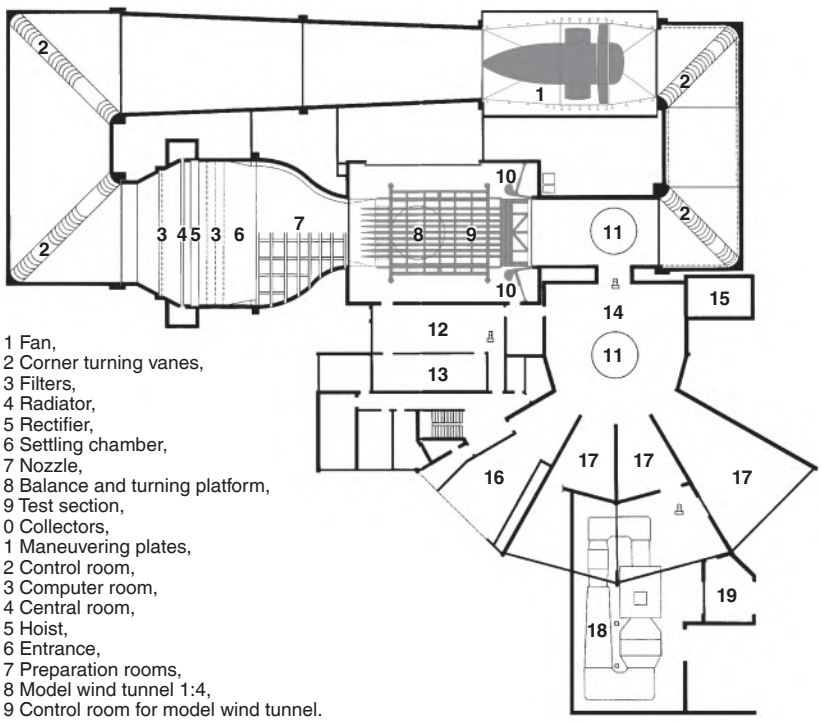
Large-scale heat exchangers are used to set a temperature range, e.g. from –40 °C to +70 °C at a low control tolerance limit of approx. ± 1 K.

The vehicle is positioned on dynamometer rollers and driven at the required load condition or load cycle. Wind speed and roller speed must be exactly synchronized, even at low speeds. If necessary, uphill or downhill gradients can be simulated to introduce realistic road factors to the vehicle operating cycle.

There is also the option of regulating the air humidity or simulating solar radiation (using a bank of lamps).

The extensive aerodynamic challenges in vehicle design are not all covered by enhanced test-bench simulations as described above. In addition to experimental testing, manufacturers are therefore making increasing use of CFD models (Computational Fluid Dynamics). They allow preliminary decisions to be made to reduce the test workload.

Figure 19: Vehicle wind tunnel (example of the Göttingen type from Dr. Ing. h.c. F. Porsche AG)



1 Fan,
2 Corner turning vanes,
3 Filters,
4 Radiator,
5 Rectifier,
6 Settling chamber,
7 Nozzle,
8 Balance and turning platform,
9 Test section,
10 Collectors,
11 Maneuvering plates,
12 Control room,
13 Computer room,
14 Central room,
15 Hoist,
16 Entrance,
17 Preparation rooms,
18 Model wind tunnel 1:4,
19 Control room for model wind tunnel.

Vehicle acoustics

Vehicle acoustics deals with all vibrational excitations and their transmission in the vehicle and in airborne noise. The primary sources here are the engine, excitation of the tires and chassis due to the texture and irregularities of the road, as well as excitation due to the airflow around the vehicle with increasing driving speed.

Table 1: Limits in dB(A) for noise emission from motor vehicles as per [1]

Vehicle category	since October of 1995 dB(A)
Passenger cars	
With spark-ignition or diesel engine	74
– with direct-injection diesel engine	75
Trucks and buses	
Permissible total weight below 2 t	76
– with direct-injection diesel engine	77
Buses	
Permissible total weight 2 t to 3.5 t	76
– with direct-injection diesel engine	77
Permissible total weight above 3.5 t:	
– engine power output up to 150 kW	78
– engine power output above 150 kW	80
Trucks	
Permissible total weight 2 t to 3.5 t	76
– with direct-injection diesel engine	77
Permissible total weight above 3.5 t (German Road Traffic Licensing Regulations: above 2.8 t):	
– engine power output up to 75 kW	77
– engine power output up to 150 kW	78
– engine power output above 150 kW	80
Off-road and four-wheel drive vehicles These vehicles are governed by higher additional limit values for engine brake and compressed-air noises.	

Legal requirements

Noise emission from vehicles and its testing

Test procedures used to monitor compliance with legal requirements are concerned exclusively with exterior-noise levels. The EC Directive 70/157/EEC enacted in 1970, together with its last revision 2007/34/EC [1], defines measurement procedures for noise emissions from stationary and moving vehicles, as well as limit values for noise emissions from moving vehicles (Table 1) for various vehicle categories. On account of the inadequate effectiveness of these statutory provisions in real traffic, the measurement method has been revised with the aim of better reproducing urban traffic in the test procedure.

The new EU noise legislation – Regulation 540/2014/EC [2] – has been in force since April 2014 and its application has been mandatory for new homologations of vehicles since July 2016. The UN Regulation UN R51.03 [3] applied in parallel in the EU has been in force since January 2016 and follows the same timeline as the EU Regulation. For the old legislation there are transitional provisions until July 2022 (until July 2023 for N2 category vehicles).

In addition, tread noise requirements were defined in the regulation 661/2009/EC [4]. Depending on the tires, the test area is 70 or 80 km/h.

Measuring noise emissions from moving passenger cars and trucks up to a permissible total weight of 3.5 t

The vehicle approaches line AA, which is located up to 10 m from the microphone plane, at a constant velocity of 50 km/h (Figure 1). After the vehicle reaches line AA, it continues under full acceleration as far as line BB (placed 10 m behind the microphone plane), which serves as the end of the test section. Passenger cars with manual transmissions and a maximum of four forward gears are tested in second gear. Consecutive readings in second and third gear are employed for

vehicles with more than four forward gears, while sports cars are tested in third gear in accordance with the definition in the Directive. Vehicles with automatic transmissions are tested in the D position.

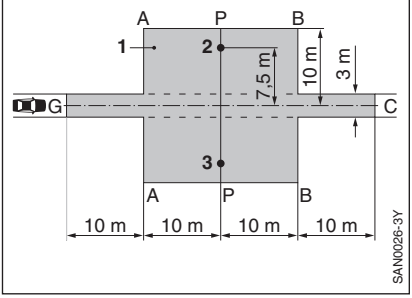
The noise-emissions level is defined as the maximum occurring sound level which is measured on the left and right sides of the vehicle at a distance of 7.5 m from the center of the lane. In a test in two gears, the noise-emissions level is the arithmetic mean from the measurement of both gears.

Measuring noise emissions from moving trucks starting from a permissible total weight of 3.5 t

The vehicle approaches line AA, which is located up to 10 m from the microphone plane, at a constant velocity (Figure 1). After the truck reaches line AA, it continues at full acceleration as far as line BB (also placed 10 m from the microphone plane), which serves as the end of the test section. The driving speed adopted is dependent on the gear tested and the rated engine speed. The choice of gears selected is based on the fact that at least the rated speed must be reached within the test section, however, the vehicle must pass through the test section without reaching the engine-speed limitation. The noise-emissions level is defined as the maximum occurring sound level from all the measurements.

Figure 1: Test layout for driving-noise measurement

1 Road surface as specified by ISO 10844 [5],
2 Left microphone,
3 Right microphone.



Measuring noise emissions from stationary vehicles

Following noise approval of a vehicle, measurement of noise emissions from stationary vehicles makes it possible to determine a reference noise value for the noise emissions of the vehicle in traffic when in “proper” condition. This makes it easier for the vehicle’s noise emissions to be checked, for example, by the police or regulatory official. Noise emission from stationary vehicles is measured at a distance of 50 cm from the exhaust outlet at a horizontal angle of 45° ± 10° relative to the direction of exhaust flow. During the measurement, the engine has to be brought up to a speed dependent on its nominal speed. Upon reaching this test speed, it is to be maintained for 3 sec. and then the throttle response switch is to be quickly moved to the idle position. The maximum A-weighted sound pressure level determined during this measurement is entered in the vehicle documentation with the suffix P´ (to distinguish it from data of earlier test procedures).

A maximum deviation of 5 dB from the entered reference value is permitted for tests in traffic.

Minimum noise from vehicles

When quiet electric motors began being widely used as a drive source for vehicles, it became a problem for pedestrians: They have great difficulty recognizing these vehicles in traffic in a timely manner. Investigations have shown that this is particularly the case in the low-speed range up to about 20 to 30 km/h. To reduce the danger of collisions between vehicles and pedestrians, work is currently being done to specify an international standard for minimum noise emissions. These provisions are mandatory from July 2019 for the new homologation of vehicles and mandatory from July 2021 for all newly approved vehicles.

Development work on vehicle acoustics

Measuring equipment for acoustics
The measuring techniques listed in the following are used to measure and evaluate not only the exterior noise level, but also the interior noise level and in particular the various vehicle vibrations [6]:
– Sound-pressure recording is performed with capacitor microphones, e.g. using sound-level meters, in dB(A).
– The dummy-head measurement technique enables noise to be recorded by microphones built into the ears of an artificial head. Artificial- or dummy-head technology provides the opportunity of making realistic recordings, so that they can be compared with other similar recordings at a different place and time.
– Measuring rooms for standard sound measurements are generally equipped with highly sound-absorbent walls. Depending on the objective, these rooms can be equipped with chassis dynamometers, so that a vehicle can be evaluated even while in driving mode.
– For vibrations and structure-borne noise acceleration, sensors (mass partly under 1 g) are used which operate, for example, according to the piezoelectric principle. Laser vibrometers are used for rapid, non-contact measurements based on the Doppler principle.
– Position measuring devices are used, for example, to record movement of the shock absorbers and chassis relative to the body.
– Vibration test benches provide the option of measuring the various cases of vibrational excitation in the overall vehicle, the body, or individual components. Hydropulsers permit excitation of low frequencies smaller than 150 Hz with large vibration paths for simulating road excitation. Shakers, on the other hand, permit high frequencies up to 2 kHz for simulating engine excitation.
– Ride comfort segments simulate various kinds of road excitation, ranging from different kinds of asphalt to manhole covers or concrete transverse joints.
– Reverberation rooms enable the sound insulation of individual materials or components to be determined, for example, for doors or whole bulkheads in the engine compartment.

Calculating methods in acoustics
Vibrations and oscillations
Natural-vibration calculations are performed using the finite-element method (FEM). Model adjustment with experimental modal analysis or modeling of forces acting during operation enables the calculation of real operational vibration shapes. In this way, structures can be optimized in early design phases with regard to their vibration response and sound radiation.

Airborne noise and fluid-borne noise
Sound-field calculations, for example of cabinet radiation or in cavities, are carried out using the finite-element method (FEM) or boundary-element method (BEM) ([7]).

Ride comfort simulation
Effects of vibration from the road on the chassis and body can be simulated by using rigid body models in combination with the road excitation. What is interesting here is that it is also possible to connect the drive unit via the engine mount and transmission mount. Such models enable variable coordination of the chassis mount and power train mount, typically in the frequency range up to 50 Hz.

Sources of noise

Engine and transmission acoustics
Engine noises are caused primarily by the mechanical movement of components. Even the combustion process in the cylinders causes excitations of the engine structure. In all cases, the vibrations are, on the one hand, directed through the structure and, on the other hand, radiated as airborne noise. In the process, noises can also occur dominantly at a clear distance from the excitation location. This is determined by the sensitivity of the membrane surfaces and the natural-vibration response of the local structure.

Important sources of noise on the engine are the crankshaft drive, the camshaft drive, the crankcase, the oil pan, the cylinder-head cover, the auxiliary power take-offs (e.g. alternator and water pump), and the belt drives. All of these elements together shape the basic acoustic character of the unit (engine/transmission combination).

The objective of the acoustician is to achieve harmonization of the vibration and noise behavior in the power train by means of suitable technical modifications. Some of the important influencing factors to take into account include the following:
– Tolerances, clearances and structures,
– Thermal expansion in clearances,
– Avoiding resonances,
– Choice of materials,
– Vibration shapes of the engine/transmission unit,
– Muffler design.

Here, most of these are in a conflict of interests with other development disciplines, particularly space design (package), energy management and driving dynamics. Thus, in a four-cylinder engine, the often dominant second engine order, which is frequently perceived as humming, can be significantly reduced with balancer shafts and very rigid units. But this comes at the expense of additional weight, more complicated structures and a considerable dip in performance.

Engine and transmission acoustics also has the task of largely reducing the transmission of rigid-body sound vibrations to the body and the airborne noise. Starting points for this are the design of the power train mount with respect to insulation, low-vibration designs, short, compact connections of auxiliary units, and prevention of large, smooth vibration surfaces by means of ribbing or additional damping materials.

The engine mount deserves special attention. Stiff assembly bearings improve driving dynamics, but convey to the driver a very raw engine noise and increase the vibration input into the body. If the bearings are too soft, it means the engine can move large distances, which can lead to unpleasant bucking motion of the whole vehicle. Highly insulating, hydraulically damped elements can be used to bring about a certain softness with good vibration comfort for the whole vehicle.

Gas exchange noise
(exhaust and intake system)
The energy transmitted by the combustion processes to the inducted air is output in time with the ignition processes and the valve movements via the exhaust-gas system to the surroundings again. The design of the exhaust manifolds, the exhaust pipes and the mufflers decisively affects which frequency components are radiated with which levels. Here, the sound output, not only at the orifice but also at the surface of all the components, is important. The noise dynamics of this acoustic source are enormous, easily reaching 15 dB and higher when the level difference is considered between idle and full load. On the basis of these dynamics, this source is ideally suitable for sound designs.

Chassis acoustics
Excitation from the road introduces vibrations into the chassis. These are strongly attenuated at first by the damping of the tires, but vibrations can still be transmitted in the chassis and into the body. Moreover, the road can incite natural resonance in the chassis, which can reach the passenger compartment in the form of unpleasant droning if the body has a certain sensitivity. Chassis acoustics has the task of finding an optimum acoustical configuration of the chassis without impairing the dynamic handling characteristics of the vehicle.

Body acoustics
The body is the interface between the chassis, power train and mounts of many auxiliary drives. These include, for example, fans, servomotors and even loudspeakers. The body, however, is also a point of attack for airflow forces. It surrounds the driver and passengers and protects them from external influences. But it also transmits vibrations and introduces them to the passenger compartment, where the occupants can both feel and hear them.

Body acoustics deals specifically with minimizing the transmission of sound and preventing airborne noise excitations, which, along with resonance in the passenger compartment, can lead to unpleasant droning noise. It also deals with optimizing the natural-oscillation characteristics of the body, because the body's eigenmodes – such as the torsion of the basic body or the first bending mode of the front end – can lead to shaking of the vehicle, which feels like when the vehicle is very fluttery on the road.

High rigidity of the attachment points for the chassis and engine provide for low excitation, sealing the byways reduces the conduction of sound within the body, and damping materials provide for sufficient insulation.

Body acoustics also includes aeroacoustics, in which optimized transitions at vehicle edges and flows around components should prevent turbulent flows from developing. These can make themselves apparent as unpleasant interior and exterior noise at high speeds.

Ride comfort
Ride comfort concerns all of the ride-related components that the driver and passengers touch. In particular these include the steering wheel, shift lever, pedals, and of course the seat. Excitations can come from the road, body, or engine and are in the frequency range under 50 Hz. The objective is to minimize all vibrations.

Dynamic strength
All vehicle components are subject to vibrational excitation, and therefore to dynamic loads. To ensure that all components have a sufficient dynamic load capacity over a long time period, they have to be investigated with respect to their dynamic strength. To do so, individual components as well as entire assemblies are subjected to continuous loading on test benches by electrodynamic shakers or hydraulic punches. The excitation profiles are gained from load measurements from real driving profiles. If necessary, the components can also be tested with climatic adaptations, such as low or extremely high temperatures. Exhaust systems can, for example, have hot gases pumped through them to simulate the engine exhaust gases. These tests provide valuable indications for optimizing details, as a result of which materials, flanges or ribs are often optimized again.

Sound design

Definition of sound design
Sound design is the active and specific matching of a product to a desired noise behavior. The aim is to lend the product a typical "sound" expected by the customer and corresponding to the brand identity and to arouse quite specific emotional associations with the product.

The most well-known field is the automobile sector, where both customers and the press show great interest in the result of the sound development of products. However, intensive work is also being carried out on the sound of products in many other fields, such as, for example, household appliances (vacuum cleaners and washing machines) and even foods (the "crunch" of potato chips).

Terms other than sound design are also commonly encountered. Sound engineering is, as a rule, the concrete technical realization on the product, i.e. the specific technical implementation of a target sound elaborated in sound design.Sound cleaning is the suppression of unwanted noise to achieve an optimally neutral noise behavior. This is often carried out for luxury vehicles as a starting point for developing a discreet sound.

Sound at the vehicle
The steps for implementing sound at a product, are generally speaking, always the same. For clarification purposes, this is explained by reference to the automobile sector.

Sound is the audible feedback of the vehicle at a specific operating state arrived at by the driver using the accelerator pedal and choosing a gear. Sound is accordingly described by the engine's operating point with regard to engine revs and load, as well as their time variance – i.e. acceleration, steady-state driving or deceleration. Starting and idle noises are added.

Depending on the design, the subjective perception of the throttle response is "spontaneous" or "sluggish", regardless of the real physically measurable acceleration.

A distinction is made between a vehicle's interior and exterior noises. Exterior noise is shaped by the airborne noise radiation of all the components, primarily by the engine and the exhaust-gas and air-intake systems and their orifices. In addition to these, there is disruptive noise from rolling and the wind. The acoustic conditions in the passenger compartment are additionally influenced by the structure-borne noise and the acoustic transfer paths through the body.

Sound quality
Sound quality refers to discrete components that can be actively moved or changed, such as switches, levers, servomotors, or flaps. Here, the actuating noise is intended to underline the quality and the soundness of the product. Opening and closing doors and hoods also has to sound good and must not be jarring. Moreover, sound quality also includes prevention of rattling, grinding and squeaking noises produced by components touching and rubbing against each other.

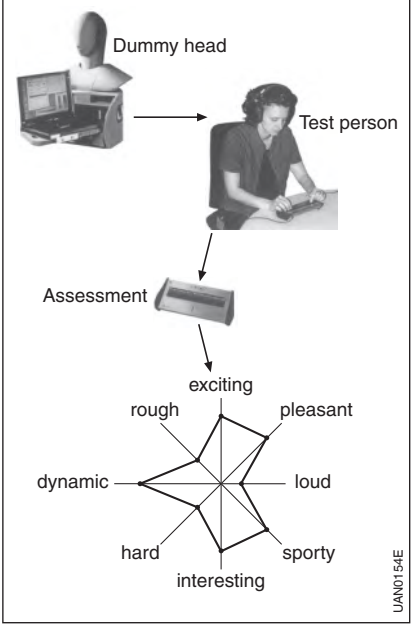
Assessment methods
Psychoacoustics
Sound cannot be assessed with the standard assessment method based on loudness, roughness or tonality. When the sound of a vehicle is mentioned, for the most part terms such as "sporty", "sonorous", "aggressive" or "inspiring" are used. These must be translated into a technical language so that the development engineers know at a later stage what spectral content, what dynamics and what time variances are responsible for them. For this reason, these subjective factors are being increasingly brought into play and

structured in order to establish assessment criteria from them. In the practical course of a psychoacoustic assessment (Figure 2), sound samples are assessed according to antagonistic value pairs (such as good – bad, loud – quiet, weak – strong) so as to arrive at a characterization of the sound.

Virtual prototyping
Analyses and measurements conducted on components can be used to make predictions about sound behavior and the leeway for making modifications. Vehicle measurements are also made using artificial-head technology. Variations can then be simulated in a narrow range with modern computer technology, always with a view toward whether a "virtually created sound" can also be technically realized.

Now the virtual sounds processed using computer technology can be evaluated by test subjects with respect to previously developed evaluation criteria. From this, it is possible to derive assessor preferences, which are now, however, technically stored and provide indications as to which frequency elements are necessary or offending and which components come into consideration.

Figure 2: Psychoacoustic assessment



Active sound
Of course, missing sound elements can also be provided with active speaker systems inside and outside the vehicle. This technology is currently gaining ever more importance, because electric vehicles and hybrid vehicles in electric driving mode are frequently evaluated as being too quiet. Significant improvements can be achieved here by adding electronic sound.These provisions are mandatory from July 2019 for the new homologation of vehicles and mandatory from July 2021 for all newly approved vehicles (see Minimum noise from vehicles).

References
[1] Commission Directive 2007/34/EC of 14 June 2007 amending, for the purposes of its adaptation to technical progress, Council Directive 70/157/EEC concerning the permissible sound level and the exhaust system of motor vehicles.
[2] Regulation (EU) No. 540/2014 of the European Parliament and of the Council of 16 April 2014 on the sound level of motor vehicles and of replacement silencing systems, and amending Directive 2007/46/EC and repealing Directive 70/157/EEC.
[3] UN R51-03: Uniform Provisions Concerning the Approval of Motor Vehicles Having a Least Four Wheels with Regard to their Sound Emissions.
[4] Regulation (EC) No 661/2009 of the European Parliament and of the Council of 13 July 2009 concerning type-approval requirements for the general safety of motor vehicles, their trailers and systems, components and separate technical units intended therefor.
[5] ISO 10844:2011: Acoustics – Specification of test tracks for measuring noise emitted by road vehicles and their tyres.
[6] Klaus Genuit: Sound-Engineering im Automobilbereich. 1st ed., Springer, 2010.
[7] L.C. Wrobel, M.H. Aliabadi:
The Boundary Element Method. 1st ed., Wiley, 2002.

Lubricants

Terms and definitions

Lubricants
Lubricants provide mutual insulation for components in a state of relative motion. The lubricant's function is to prevent direct contact between these components, thereby reducing wear and minimizing friction. Lubricants serve as coolants, friction-surface sealants, and corrosion inhibitors, and can also reduce running noise. Lubricants may be solid, consistent, liquid, or gaseous in form. Specific lubricants are selected with reference to design characteristics, materials combinations, the ambient conditions, and the stress factors encountered at the friction surface ([1], [2], [3]).

Additives
Additives are substances mixed into the lubricant in order to improve specific properties. These substances modify either the lubricant's physical characteristics (e.g. viscosity-index improvers, pour-point depressors) or its chemical properties (e.g. oxidation inhibitors, corrosion inhibitors). In addition, the properties of the friction surfaces themselves can be modified by additives which change the friction characteristics (friction modifiers), protect against wear (anti-wear agents), or provide protection against scoring and seizure (extreme-pressure additives). Great care must be exercised in order to ensure that the additives are correctly matched with each other and with the base lubricant.

ATF
ATF (automatic transmission fluids) are special-purpose lubricants specifically formulated to meet stringent requirements for operation in automatic transmissions.

Ash
Ash is the mineral residue which remains after oxide and sulfate incineration (DIN 51575, [4]).

Bleeding
Bleeding is the separation of base oil (lube oil) and the thickener in a lubricating grease (oil separation, DIN 51817 [5]). During this process the lube oil is released to the lubricating surroundings.

Bingham bodies
Bingham bodies are materials whose flow characteristics differ from those of Newtonian fluids (see Rheology, Newtonian fluids).

Fire point and flash point
The lowest temperature (referred to 1,013 hPa) at which a gaseous mineral product initially flashes is known as the flash point. The fire point is the temperature at which it continues to burn for at least 5 s (DIN EN ISO 2592 [6]).

Cloud point
The cloud point is the temperature at which mineral oil becomes opaque due to the formation of paraffin crystals or precipitation of other solids (DIN ISO 23015 [7]).

Detergents
These additives prevent paint- and carbon-like deposits on hot components (e.g. pistons). They serve as cleaning agents which like a tenside contain both hydrophilic and hydrophobic areas.

Dispersants
Dispersants prevent (disperse) sludging and sedimentation, particularly at low temperatures (solid substances are held finely distributed in suspension).

EP lubricants
See High-pressure lubricants (EP, Extreme Pressure).

Flow pressure
The flow pressure is the pressure required to press a consistent lubricant through a standardized test nozzle (Kesternich method, DIN 51805 [8]). The flow pressure is an index of a lubricant's starting flow characteristics, particularly at low temperatures.

Yield point
The yield point is the shear stress at which a substance begins to flow. Above the yield point, the rheological characteristics of plastic substances are the same as those of liquids (DIN 1342-3 [9]).

Friction modifiers
Friction modifiers are polar lubricant additives which reduce friction in the mixed-friction range (see Mixed friction) and increase bearing capacity after adsorption on the surface of the metal. They also inhibit stick-slip behavior.

Gel-type greases
Gel-type greases are lubricants with inorganic gelling agents (e.g. Bentonites, silica gels).

Anti-friction coatings
Anti-friction coatings are solid lubricant combinations which a binding agent holds in place on the friction surfaces (AFC, anti-friction coating).

Graphite
Graphite is a solid lubricant with layer-lattice structure. Graphite provides excellent lubrication when combined with water (e.g. high atmospheric humidity) and in carbon-dioxide atmospheres or when combined with oils. It does not inhibit friction in a vacuum.

Borderline pumping temperature
The borderline pumping temperature is the application limit for oil throughput through engines after cold starting. Below the borderline pumping temperature oil can no longer flow of its own accord to the oil pump, and thus a sufficient supply of oil can no longer be guaranteed.

High-pressure lubricants
High-pressure lubricants (EP, Extreme Pressure) contain additives to enhance load-bearing capacity, to reduce wear, and to reduce scoring (generally provide good performance in steel-to-steel and steel-to-ceramic applications).

Hydro-crack oils
Hydro-crack oils are refined mineral oils with increased viscosity index (VI 130 to 140). They are produced by hydrogenating mineral oils and are thus more thermally stable and have improved viscosity-temperature characteristics.

Induction period
The induction period is the time which elapses before substantial changes occur in a lubricant (e.g. aging of an oil containing an oxidation inhibitor).

Inhibitors
Inhibitors are active system-protecting ingredients which slow down the chemical decomposition of lubricants (e.g. due to oxidation) or protect against corrosion (corrosion inhibitors).

Low-temperature sludge
Low-temperature sludges are the products of oil degradation which form in the engine crankcase due to incomplete combustion and condensate at low engine load. Low-temperature sludge increases wear and can cause engine damage. Modern high-quality engine oils inhibit its formation.

Cold-starting reliability
Cold-starting reliability is a temperature value and based on experience is approx. 5 °C above the borderline pumping temperature.

Consistency
Consistency is a measure of the ease with which lubricating greases and pastes can be deformed (DIN ISO 2137 [10]). A standardized cone is dropped into a flattened grease sample and then the penetration depth is measured in 1/10 mm (Figure 2, see also Penetration and Table 5).

Doped lubricants
Doped lubricants contain additives for improving specific properties (e.g. aging stability, wear protection, corrosion protection, viscosity-temperature characteristics).

High-lubricity oils
High-lubricity oils are lubricating oils with multigrade properties, low cold viscosity, and special anti-friction additives. Extremely low engine friction under all operating conditions reduces fuel consumption.

Longlife engine oils
Longlife engine oils are oils for significantly extended oil-change intervals.

Multigrade oils
Multigrade oils are engine and transmission oils with good resistance to viscosity-temperature change (high viscosity index VI, Figure 1, see Viscosity). They often contain polymers which change their solid structure as a function of temperature. When the oil is cold, the molecules in the oil are "bunched together". As the temperature increases, the molecules stretch and thereby increase the friction between the particles. The viscosity of the oil increases, thereby guaranteeing a stable lubricating film even at high temperatures. Multigrade oils reduce friction and wear, and during cold starting ensure that all the engine components are quickly lubricated. These oils are formulated for year-round use in motor vehicles; their viscosity ratings extend through several SAE grades.

Metal soaps
Metal soaps are reaction products from metals or from their compounds with fatty acids. They are used as thickeners for grease and as friction modifiers.

Mineral oils
Mineral oils are distillates or raffinates produced from petroleum or coal. They consist of numerous hydrocarbons in various chemical compositions. Classification is according to the predominant component: paraffin-based oils (chain-shaped saturated hydrocarbons), naphthene-based oils (closed-chain saturated hydrocarbons, generally with five or six carbon atoms per ring) or aromatic oils (e.g. alkylbenzene). These substances are distinguished by major variations in their respective chemical and physical properties.

Molybdenum disulfide
Molybdenum disulfide (MoS_2) is a solid lubricant with layer-lattice structure. Only low cohesive forces are present between the individual layers, so their mutual displacement is characterized by relatively low shear forces. A reduction in friction is only obtained when MoS_2 is applied in suitable form to the surface of the metal (e.g. in combination with a binder such as MoS_2 anti-friction coating).

Figure 1: Viscosity/temperature curves for single- and multigrade engine oils

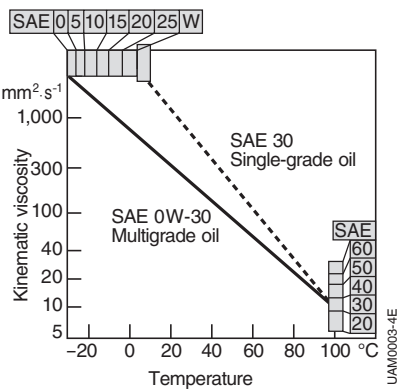
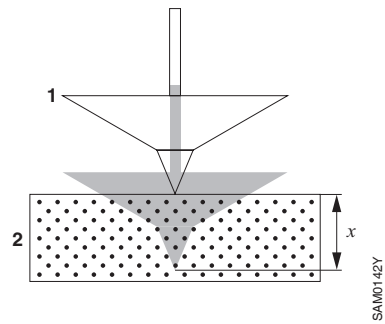


Figure 2: Determining cone penetration
1 Standardized cone,
2 Consistent lubricant at defined temperature.
x Penetration depth.



Penetration
Penetration denotes the depth (in 10^{-1} mm) to which a standardized cone penetrates into a consistent lubricant within a defined period and at a specified temperature (Figure 2). The larger the number, the softer the lubricant (DIN ISO 2137 [10]).

Polar substances
Dipolar molecules are easily adsorbed onto metal surfaces. They enhance adhesion and bearing capacity, thus reducing friction and wear. This category includes, for example, diester oils, ethers, polyglycols, and fatty acids.

Pour point
Pour point is the lowest temperature at which an oil continues to flow when cooled under defined conditions (DIN ISO 3016 [11]).

PTFE
PTFE (polytetrafluoroethylene, Teflon) is a thermoplastic with outstanding properties as a solid lubricant, particularly at very low sliding velocities (< 0.1 m/s). PTFE only becomes brittle below approx. -270 °C. The upper service temperature for use is approx. 260 °C. Above this level, it decomposes with toxic cleavage products.

Rheology
Rheology is the science dealing with the flow characteristics of materials. These are generally represented in the shape of flow curves (Figure 3). In the diagram shear stress τ is plotted against shear rate $\dot{\gamma}$.

Shear stress and shear rate
Shear stress is defined as the shear force per shear area, shear rate as the quotient of velocity v and gap height h (Figure 4).

Shear stress $\tau = F/A$ (in N/m^2 = Pa),
F shear force, A area vs. shear rate.

Shear rate $\dot{\gamma} = v/y$ (in s^{-1}),
v velocity, y thickness of lubricating film.

The flow curves are usually measured in a plate-cone measuring system. The sample to be measured is located between the plate and the cone; the torque is proportional to the shear stress and the rotational speed proportional to the shear rate.

Shear stress τ can be illustrated using the two-plate model (Figure 4). The fluid is located between the two plates. The bottom plate is fixed, the top plate moves with shear area A by means of shear force F.

Shear viscosity
Shear viscosity η is in liquids of ideal viscosity and constant temperature the ratio of shear stress τ to shear rate $\dot{\gamma}$:

$$\eta = \frac{\tau}{\dot{\gamma}}.$$

The term “dynamic viscosity” is also used for η. The former unit centiPoise (cP) is equal to the unit mPa · s.

Figure 3: Flow curves
1 Rheopexic, 2 Thixotropic, 3 Newtonian, 4 Plastic, 5 Dilatant, 6 Intrinsically viscous, 7 Yield point.

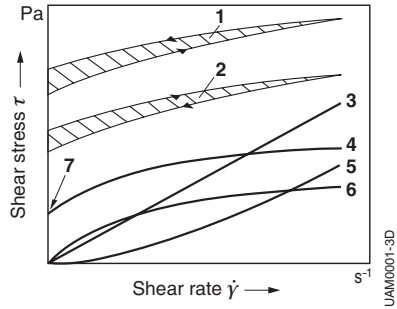
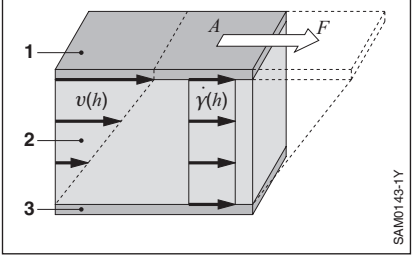


Figure 4: Velocity distribution and shear rate in the lubricating gap
1 Moving plate, 2 Fluid, 3 Fixed plate.
A Shear surface, F Shear force, h Height, v Velocity, $\dot{\gamma}$ Shear rate.



Midpoint viscosity
Each viscosity grade in accordance with DIN ISO 3448 [12] is defined by a viscosity range at a temperature of 40 °C (see Table 1). The permissible limit is ±10 %. Thus, for example, viscosity grade ISO VG 10 has the limits 9 and 11 mm^2/s at a midpoint viscosity of 10 mm^2/s.

HTHS viscosity
The HTHS viscosity (High Temperature High Shear) specifies the shear viscosity at a temperature of 150 °C and a shear rate of $10^6\ s^{-1}$.

Kinematic viscosity
The kinematic viscosity ν is usually determined in vertical, narrow glass tubes. The time the fluid needs at a defined temperature to flow through a specific section of the tube is measured. If density ρ is known, it is also possible to calculate the shear viscosity η from it.

$\nu = \frac{\eta}{\rho}$, (ν in mm^2/s; ρ in kg/m^3).

The former unit "centiStokes" (cSt) is equal to the unit mm^2/s.

Newtonian fluids
Newtonian fluids display a linear relationship between shear stress τ and shear rate $\dot{\gamma}$ in the shape of a straight line through zero, with the slope increasing as a function of viscosity (Figure 3).

All materials not characterized by this kind of flow behavior are classified as non-Newtonian fluids.

Intrinsically viscous flow behavior
Intrinsically viscous flow behavior is the capacity of a fluid to show decreasing viscosities with increasing shear rate (Figure 3, e.g. liquid greases, multigrade oils with viscosity-index improvers).

Dilatant flow behavior
Dilatant flow behavior is the increase in viscosity with increasing shear rate (Figure 3).

Plastic flow behavior
A plastic substance is a substance whose rheological behavior is characterized by a yield point (according to DIN 1342-1 [13]). Plastic flow behavior is the formability of an intrinsically viscous fluid supplemented by yield point (Figure 3, e.g. lubricating greases).

Thixotropy
Thixotropy is a characteristic of those non-Newtonian fluids that display an decrease in viscosity proportional to shear time, and only gradually recover their original viscosity once shearing ceases (Figure 3).

Rheopexy
Rheopexy is a characteristic of those non-Newtonian fluids that display an increase in viscosity proportional to shear time, and only gradually recover their original viscosity once shearing ceases (Figure 3).

Table 1: Viscosity grades for industrial lubricating oils acc. to ISO 3448 [12]

ISO viscosity grade	Medium viscosity at 40 °C in mm^2/s	Limits of kinematic viscosity at 40 °C in mm^2/s	
		min.	max.
ISO VG 2	2.2	1.98	2.42
ISO VG 3	3.2	2.88	3.52
ISO VG 5	4.6	4.14	5.06
ISO VG 7	6.8	6.12	7.48
ISO VG 10	10	9.00	11.0
ISO VG 15	15	13.5	16.5
ISO VG 22	22	19.8	24.2
ISO VG 32	32	28.8	35.2
ISO VG 46	46	41.4	50.6
ISO VG 68	68	61.2	74.8
ISO VG 100	100	90.0	110
ISO VG 150	150	135	165
ISO VG 220	220	198	242
ISO VG 320	320	288	352
ISO VG 460	460	414	506
ISO VG 680	680	612	748
ISO VG 1000	1,000	900	1,100
ISO VG 1500	1,500	1,350	1,650
ISO VG 2200	2,200	1,980	2,420
ISO VG 3200	3,200	2,880	3,520

Dropping point
The dropping point is the temperature at which a lubricating grease attains a specified viscosity under specified test conditions (DIN ISO 2176 [14]).

Stribeck curve
The Stribeck curve portrays friction levels between two liquid- or grease-lubricated tribological systems separated by a narrowing gap (e.g. lubricated plain or ball bearings) as a function of sliding velocity (Figure 5). As the relative speed increases, so too does the hydrodynamic pressure in friction contact.

Solid-body friction
In the case of solid-body friction, the height of the lubricating film is lower than that of the roughness protrusions in the material's surface. This causes wear.

Mixed friction
In the case of mixed friction, the height of the lubricating film is approximately equal to that of the roughness protrusions. This still means increased wear because the roughness protrusions are in contact with each other.

Hydrodynamics
In the case of hydrodynamics, there is complete separation between the primary body and the opposed body, e.g. when aquaplaning (virtually wear-free condition).

Synthetic oil
Synthetic oil is manufactured in a process of chemical synthesis from smaller molecules. For example, synthetic hydrocarbons in the form of poly-α-olefins, which are manufactured by the polymerization and subsequent hydrogenation of ethylene. Further synthetic oils are polyglycols, diester oils, silicone oils, and perfluoropolyether oils.

Viscosity
Viscosity is a measure of the internal friction of substances (DIN 1342 [13], DIN EN ISO 3104 [15]). It is caused by the resistance (internal friction) with which the substance's molecules oppose displacement forces (see Rheology). As the temperature decreases, the viscosity increases and the oil becomes more viscous. At low temperatures the viscosity must not be too high so that the oil in the bearings does not exert too excessive a resistance on the rotational motion of the engine or transmission. At high temperatures the oil must still be sufficiently viscous to maintain the lubricating film.

Viscosity index
The viscosity index (VI) is a mathematically derived number expressing the change in a mineral-oil product's viscosity relative to its temperature. The greater the viscosity index, the lower the effect of temperature on the viscosity (DIN ISO 2909 [16]).

Viscosity grades
Oils are classified within specific viscosity ranges in viscosity grades:
– ISO viscosity grades (DIN ISO 3448 [12]): see Table 1.
– SAE viscosity grades (SAE J300 [17], SAE J306 [18]): see Table 2 and Table 3.

Worked penetration
Worked penetration is the penetration of a grease sample after it has been processed in a grease kneader (DIN ISO 2137 [19]).

Figure 5: Stribeck curve
R Surface roughness, F_N Normal force,
d Distance between primary and opposed body.
Range a: Solid-body friction, high wear.
Range b: Mixed friction, moderate wear.
Range c: Hydrodynamics, no abrasive wear.

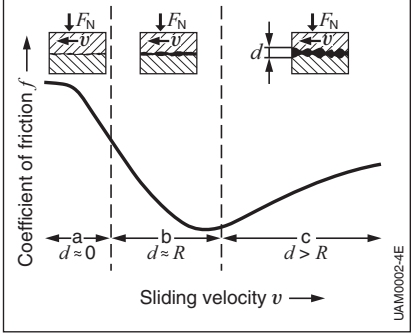


Table 2: Extract from SAE viscosity grades for transmission oils (SAE-J306 revision June 2005)

SAE viscosity grade	Maximum temperature [°C] for dynamic viscosity at 150,000 mPa·s (ASTM D 2983) [20])	Kinematic viscosity [mm²/s] at 100 °C (ASTM D 445 [21])	
		min.	max.
70 W	–55	4.1	–
75 W	–40	4.1	–
80 W	–26	7.0	–
85 W	–12	11.0	–
80	–	7.0	<11.0
85	–	11.0	<13.5
90	–	13.5	<18.5
110	–	18.5	<24.0
140	–	24.0	<32.5
190	–	32.5	<41.0
250	–	41.0	–

Engine oils

Engine oils are employed primarily to lubricate contiguous components in relative motion within the internal-combustion engine. The oil also removes heat generated by friction, carries abraded particles away from the friction surface, washes out contaminants, holds them in suspension, and protects metals against corrosion. The most common engine oils are mineral oils treated with additives (HD oils: Heavy Duty for high load, extreme operating conditions). Higher stress-resistance requirements combined with extended oil-change intervals have led to widespread application of fully and semi-synthetic oils (mixture of synthetic oil and mineral oil), e.g. hydro-crack oils. The quality of an engine oil is determined by its origin, the refining processes used on the mineral oil (except in the case of synthetic oils) and the additive composition.

Table 3: SAE viscosity grades for engine and transmission oils (SAE J300, January 2015)

SAE viscosity grade	Viscosity (ASTM D 5293 [22]) mPa·s max.	Limit pumping viscosity (ASTM D 4684 [23]) with no yield point mPa·s max.	Kinematic viscosity (ASTM D 445 [21]) mm²/s at 100 °C min.	max.	Viscosity under high shear[1] (ASTM D 4683 [24], CEC L-36-A-90 [25], ASTM D 4741 [26] or ASTM D 5481 [27]) mPa·s at 150 °C and $\dot{\gamma} = 10^6\ s^{-1}$ min.
0 W	6,200 at –35 °C	60,000 at –40 °C	3.8	–	–
5 W	6,600 at –30 °C	60,000 at –35 °C	3.8	–	–
10 W	7,000 at –25 °C	60,000 at –30 °C	4.1	–	–
15 W	7,000 at –20 °C	60,000 at –25 °C	5.6	–	–
20 W	9,500 at –15 °C	60,000 at –20 °C	5.6	–	–
25 W	13,000 at –10 °C	60,000 at –15 °C	9.3	–	–
8	–	–	4	<6.1	1.7
12	–	–	5	<7.1	2.0
16	–	–	6.1	<8.2	2.3
20	–	–	6.9	<9.3	2.6
30	–	–	9.3	<12.5	2.9
40	–	–	12.5	<16.3	3.5 (0W–40, 5W–40, 10W–40)
40	–	–	12.5	<16.3	3.7 (15W–40, 20W–40, 25W–40, 40 monogrado)
50	–	–	16.3	<21.9	3.7
60	–	–	21.9	<26.1	3.7

[1] Also called HTHS (High Temperature High Shear) viscosity.

Additives are classified according to their respective functions:
– Viscosity-index improvers
– Pour-point improvers
– Oxidation and corrosion inhibitors
– Detergent and dispersant additives
– Extreme-pressure (EP) additives
– Friction modifiers
– Anti-foaming agents

Oil is subjected to considerable thermal and mechanical stresses in the IC engine. The data on the physical properties of engine oils provide information on their application limits (SAE viscosity grades), but are not indicative of their other performance characteristics. Therefore, there are several different test procedures for evaluating engine oils:
– ACEA (Association des Constructeurs Européens de l'Automobile) standards, replaced the CCMC standards (Comité des Constructeurs d'Automobiles du Marché Commun) at the beginning of 1996
– API classification (American Petroleum Institute)
– MIL specifications (Military)
– Company specifications (e.g. ILSAC, International Lubricants Standardization and Approval Committee).

The approval criteria include the following:
– Sulfate-ash content
– Zinc content
– Engine type (diesel or spark-ignition engines, naturally aspirated or forced-induction engines)
– Load on power-transmission components and bearings
– Wear-protection properties
– Oil operating temperature (in oil pan)
– Combustion residue and chemical stress exerted on the oil by acidic combustion products
– The oil's detergent and residue-scavenging properties
– Its suitability for use with gasket and sealing materials.

ACEA (CCMC) Specifications

Engine oils for gasoline engines
– A1: Special high-lubricity oils with reduced viscosity at high temperatures and high shear for reducing viscous friction.
– A2: Conventional and high-lubricity engine oils without any restriction on viscosity grades. Higher requirements than CCMC G4 and API SH, withdrawn.
– A3: Oils of this category meet higher requirements than A2 and CCMC G4 and G5.
– A5: Improved "fuel economy" properties compared with A3 occasioned by lower viscosity including improved additive composition. Only for use in engines which have been specifically designed for this purpose.

Engine oils for passenger-car diesel engines
– B1: Corresponding to A1 for low friction losses and, consequently, reduced fuel consumption.
– B2: Conventional and high-lubricity engine oils compliant with the current minimum requirements (higher than those of CCMC PD2), withdrawn.
– B3: Exceeds B2.
– B4: Corresponds to B2, particularly suitable for VW TDI engines.
– B5: Oils surpass B3 and B4, improved "fuel economy" properties, also satisfies VW 50600 and 50601. Only for use in engines which have been specifically designed for this purpose.

Engine oils for passenger-car diesel engines with particulate filters
– C1: Since 2004, sulfate-ash content max. 0.5 %. Lowered HTHS (Ford).
– C2: Since 2004, sulfate-ash content max. 0.8 %. With HTHS ≥ 2.9 mPa · s (Peugeot).
– C3: Since 2004, sulfate-ash content max. 0.8 %. With HTHS ≥ 3.5 mPa · s (MB and BMW).
– C4: Since 2007, sulfate-ash content max. 0.5 %. With HTMS ≥ 3.5 mPa · s.

Engine oils for commercial-vehicle diesel engines
– E1: Oils for naturally aspirated and turbocharged engines with normal intervals between oil changes; withdrawn.
– E2: Derived from MB Specification Sheet 228.1. Primarily for engine designs predating the Euro II standard.
– E3: For Euro II engines, derived from MB Specification Sheet 228.3. In com-

parison with the predecessor category CCMC D5, these oils evince a significant improvement in soot dispersion capability and a much reduced tendency to thicken; withdrawn.
– E4: Diesel engines with Euro I to Euro III standards and high requirements, particularly for extended intervals between oil changes (according to manufacturer's specification). Based to a large extent on MB Specification Sheet 228.5.
– E5: For Euro III engines, reduced ash content (withdrawn).
– E6: For EGR engines (exhaust-gas recirculation) with/without diesel particulate filters and SCR-NO_x engines (Selective Catalytic Reduction, i.e. reduction of NO_x to N_2). Recommendation for engines with diesel particulate filters and operation with sulfur-free fuel, sulfate ash less than 1 % by mass.
– E7: For engines without diesel particulate filters of most EGR engines and of most SCR-NO_x engines, sulfate ash max. 2 % by mass.
– E9: For engines with/without particulate filters of most EGR and SCR-NO_x engines, sulfate ash max. 1 % by mass.

Example of a designation
According to the grade specification, a numerical code is usually supplemented.

Example: An A3/B3-04 is an engine oil for spark-ignition engines (Grade A) and diesel engines (Grade B) of quality category 3, tested accord to the ACEA classification grade issued in 2004.

API classification grades
– S Grades (Service) for gasoline engines.
– C Grades (Commercial) for diesel engines.
– SF: For engines produced in the 1980s; withdrawn.
– SG: Applicable since 1988, with more stringent sludge test, improved oxidation stability and wear protection.
– SH: Since mid-1993, corresponds to API SG quality level, but with more stringent process requirements for oil grade testing.
– SJ: Since October 1996, more tests than API SH.
– SL: Since 2001, lower oil consumption compared with SJ, lower volatility, improved engine cleanness, greater resistance to aging.
– SM: Since 2004, for gasoline and light diesel engines with improved wear protection, higher aging stability, improved pumpability, and also used oil.
– SN: Since October 2010, compared with SM improvements in piston cleanliness, sludging, and exhaust-gas treatment capability. Elastomer compatibility is also defined.
– CC: Engine oils for non-turbocharged diesel engines with low stress factors; withdrawn.
– CD: Engine oils for non-turbocharged and turbocharged diesel engines, replaced by API CF in 1994.
– CD-2: API CD requirements, plus additional requirements relevant to two-stroke diesel engines.
– CF-2: Oils with special two-stroke properties (since 1994).
– CE: Oils with CD performance characteristics, with supplementary test operation in US Mack and Cummins engines.
– CF: Replaced API CD in 1994. Specially for indirect injection, even if the sulfur content of the fuel is greater than 0.5 %.
– CF-4: As API CE, but with more stringent test procedure in single-cylinder Caterpillar turbo-diesel engines.
– CG-4: For diesel engines operating under very stringent conditions. Exceeds API CD and CE. Fuel sulfur content less than 0.5 %. Required for engines compliant with post-1994 emission-control legislation.
– CH4: Modern commercial-vehicle engine oil since 1998. Surpasses CG-4 in standards of wear, soot and viscosity. Longer oil-change intervals.
– CI-4: Since 2002 for high-speed four-stroke engines which can only still meet future exhaust-emission legislation with EGR. Suitable for sulfur content greater than 0.5 %.
– CI-4 plus: As CI-4, but with improved soot transport and higher requirements with regard to viscosity increase.
– CJ-4: Valid since 2006. For highway vehicles which must satisfy the USA

2007 emissions standards with diesel fuel (below 500 ppm sulfur content). Suitable for particulate-filter and NO_x reduction catalytic converters.
– CK-4: Improved protection for new diesel engines with among others increased aging stability, shear stability, piston cleanliness. Start 12/2016
– FA-4: For new diesel engines of the 2016/17 generation with low HTHS (2.9 to 3.2 mPa·s at 150 °C) among others to meet the required emission stipulations in the USA.

ILSAC
ILSAC (International Lubricants Standardization and Approval Committee) is a joint standard of General Motors, Daimler, the Japanese Automobile Manufacturers Association, and the American Engine Manufacturers Association.

ILSAC GF-3
In addition to API-SL, the standard requires a fuel-economy test.

ILSAC GF-4
Allocation to API-SM.

ILSAC GF-5
This standard has been in force since October 2010 with the aim of increased fuel reduction, elastomer compatibility, and improved protection of the exhaust-gas treatment systems (comparable with API-SN).

ILSAC GF-6
New specification with the aim of fuel economy for gasoline direct injection and turbocharged engines, expected around 2018. GF-6 A with conventional viscosities, GF-6B with viscosities ≤ SAE 0W16.

SAE viscosity grades
(SAE J300 [17], SAE J306 [18])
The SAE (Society of Automotive Engineers) grades are the internationally accepted standard for defining viscosity. It provides information on the temperature range in which the oils are used. The standard provides no information on the quality of the oil.

A distinction is drawn between single-grade and multigrade oils. Single-grade oils are designated for example by SAE 30. Low code numbers denote thin-bodies oils, higher code numbers denote more viscous oils.

Multigrade oils are the type in widespread use today. Two series are employed for the designation (see Table 2 and Table 3), where the letter "W" (Winter) is used to define specific cold-flow properties. These oils are also suitable for low temperatures. The viscosity grades including the letter "W" are rated according to maximum cold viscosity, maximum viscosity pumping temperature and the minimum viscosity at 100 °C.

Viscosity grades without the "W" are rated only according to viscosity at 100 °C.

Viscosity specifications at higher temperature (150 °C) and high shear stress (HTHS viscosity, see Table 3) are more practical. This value acquired particular significance with the introduction of oils with reduced high-temperature viscosity (below 3.5 mPa·s). These oils may only be used if manufacturer approvals exist.

A multigrade oil is designated for example by SAE 0W-30. This means that it has at –35 °C a viscosity of max. 6,200 mPa·s and satisfies the specification of an SAE 30 oil with a kinematic viscosity in the range of 9.3 to 12.5 mm^2/s at 100 °C and an HTHS viscosity of min. 2.9 mPa·s at 150 °C (Table 3, see also Figure 1).

Motorcycle oil
This essentially is no different from oil for automobiles. However, the stress for example with regard to shear load and specific surface pressure is if necessary higher and only the oils recommended by the manufacturer may be used.

Oil dilution
When vehicles are operated in short-distance service, unburned fuel can mix with engine oil (oil dilution, oil proliferation). This reduces the operating viscosity and increased wear can occur under mixed-friction conditions. In winter the presence of diesel fuel in engine oil can also cause paraffin-crystal precipitation, which compromises delivery (e.g. filter blockage) and can even cause deficient-lubrication conditions.

Transmission oils

Specifications for transmission oils are defined by the type of transmission and the stresses to which it is subjected throughout the entire range of operating conditions. The requirements (high pressure resistance, high viscosity stability relative to temperature, high resistance to aging, good anti-foaming properties, compatibility with gaskets and seals) can only be satisfied by lubricants treated with special additives. In contrast to engine oils, transmission oils contain no, or only minimal, detergent additives, considerably fewer basic constituents, and mostly no viscosity-index improvers. Most of them would be sheared and thus become inactive. The use of unsuitable and qualitatively inferior oils results typically in damage to bearings and gear-tooth flanks.

The viscosity must also suit the specific application. Viscosity grades for vehicular transmissions are defined in SAE J 306 [18] and SAE J300 [17] (Table 2 and Table 3). Transmission oils with roughly the same viscosity compared with engine oils are identified by higher code numbers so that they can be clearly distinguished from engine oils.

Synthesized oils are being increasingly used to meet special requirements (e.g. poly-α-olefins). Advantages over standard mineral oils include superior temperature-viscosity properties and increased aging resistance.

API classifications of transmission oils

- GL-1…GL-3: Outdated, no longer of practical significance.
- GL-4: Transmission lubricants for moderate-stressed hypoid-gear transmissions, and for transmissions which operate at extreme speeds and impact loads, high rotational speeds and low torques, or low rotational speeds and high torques.
- GL-5: Transmission lubricants for high-stressed hypoid-gear transmissions in passenger cars and in other vehicles where they are exposed to impact loads at high rotational speeds, and at high rotational speeds and low torques, or low rotational speeds and high torques.
- MT-1: For non-synchromesh manual transmissions in US trucks.

Many truck and component manufacturers have drawn up their own specifications and no longer rely on API classification grades.

Lubricants for automatic transmissions

Automatic transmissions differ from their manual-shifted counterparts in the way they transfer torque; non-positive mechanical and hydrodynamic force transfer is supplemented by friction coupling arrangements. Thus, the friction response of automatic transmission fluids (ATF) is extremely important. Applications are basically classified according to the friction characteristics.

General Motors

- Superseded grades: Type A, Suffix A, DEXRON, DEXRON B, DEXRON II C, DEXRON II D.
- DEXRON II E: Valid through 1994.
- DEXRON III F/G: Valid since 1994 and 1997. Features more stringent requirements for oxidation stability and consistency in frictional coefficient.
- DEXRON III H: Valid since 2005, but outdated since 2007.
- DEXRON VI: Valid since 2006.

Other manufacturers

(among others. Ford, MB, MAN, Mack, Scania, ZF): In accordance with fluids and lubricants specifications.

Lubricating oils

Lubricating oils comprise the components of basic oil and additive. The additives improve the properties of the basic oils, e.g. with regard to oxidation stability, corrosion protection, protection against scoring and seizure, or viscosity-temperature characteristics. In addition, system properties such as friction and wear are optimized in the desired direction.

There are a wide range of designations in the form of letters and numbers (see among others DIN 51502 [28]) for the most varied applications. For example, hydraulic fluids:

– HL: Mineral-oil-based hydraulic fluid with additives for improving corrosion protection and resistance to aging.
– HLP: As HL, but with extra additives against scoring and seizure.
– HVLP: As HLP, with extra viscosity-index improver.

Lubricating greases

Lubricating greases are thickened lubricating oils. A great advantage that greases enjoy over oil is that they do not drain from the friction surfaces. Complicated measures to seal them in place are therefore unnecessary (e.g. application in wheel bearings pumps, alternators, windshield-wiper motors, servo motors).

Composition

Table 4 provides a general overview of the components in a consistent lubricating grease as blended from three basic components – base oil, thickener, and additive.

Mineral oils are usually employed as the basic oil component, although full-synthetic oils have recently become more common as a replacement (e.g. due to more stringent requirements for aging stability, cold-flow properties, viscosity-temperature characteristics).

Thickeners are used as a binder for the base oil; metal soaps being generally employed (Figure 6). They bind the oil in a sponge-like soap structure (micelle) by means of inclusions and physical Van der Waals forces. The higher the proportion of thickener (depends on the type of thickener) in the grease, the less the penetration and the higher the NLGI grade

Table 4: Composition of lubricating greases

Whatever the friction pairing, the large number of lubricant components can be used to develop a high-performance lubricant.

Base oils	Thickeners	Additives
Mineral oils – paraffinic – naphthenic – aromatic Poly-α-olefins Alkyl aromatics Diester oils Polyhydric alcohol Silicones Phenyl ether oils Perfluoropolyethers	Metal soaps of metals Li, Na, Ca, Ba, Al Normal soaps (soaps with a carbonic acid, e.g. stearic acid) Hydroxy soaps (soaps with an additional hydroxide group, e.g. 12-hydroxystearic acid Complex soaps (e.g. Ca soaps with a short- and a long-chain carbonic acid) Polyureas PTFE PE Bentonites Silica gels	Oxidation inhibitors Fe, Cu ions, complexing agents Corrosion inhibitors High-pressure additives (EP additives) Wear-protection additives (anti-wear additives) Friction reducers (friction modifiers) Adhesion improvers Detergents, dispersants VI improvers Solid lubricants

(National Lubricating Grease Institute, US standard, see Table 5).

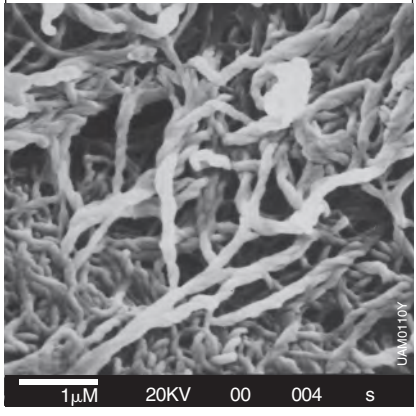
The additives serve to modify physical and chemical properties of the lubricating grease to achieve specific objectives (such as improvement of anti-oxidation properties, increased protection against scoring and seizure, and reduction of friction and wear).

Solid lubricants (e.g. MoS_2) are also added to lubricating greases (for instance for lubricating constant-velocity joints in motor vehicles).

Table 5. Consistency grades for lubricating greases (DIN 51818 [29])

NLGI grade	Worked penetration acc. to DIN ISO 2137 [19] in units of 0.1 mm
000	445…475
00	400…430
0	355…385
1	310…340
2	265…295
3	220…250
4	175…205
5	130…160
6	85 to 150 (unworked penetration)

Figure 6: Photo of lithium soap taken by a scanning electron microscope
The oil is retained between the twisted soap fibrils.



Selection

Specific lubricating greases are selected with reference to their physical characteristics and their effects on the friction surface, and to minimize interaction between the grease and the contact materials. Mutually antagonistic effects with polymers are, for example: formation of stress cracks (Figure 7), consistency changes, polymer degradation, swelling, shrinkage, and brittleness.

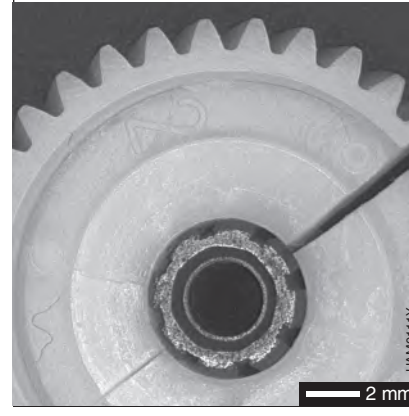
Thus, for example, mineral-oil greases and greases based on synthetic hydrocarbons should not come in contact with elastomers used together with brake fluid (polyglycol base) (e.g. substantial swelling of EPDM elastomers).

In addition, lubricating greases with varying compositions should not be mixed (changes in physical properties, grease liquefaction due to drop-point reduction).

Stress

Thermal and mechanical stresses result in chemical and/or physical changes which may have a detrimental effect on the function of the entire tribological system (Figure 8). Oxidation, for example, results in acidification, which can trigger corrosion on metal surfaces or stress cracking on some plastics. In the event of excessive thermal load, polymerization can cause the lubricant to solidify.

Figure 7: Stress cracks on a gearwheel made of polyoximethylene (POM), caused by poly-α-olefin (PAO)

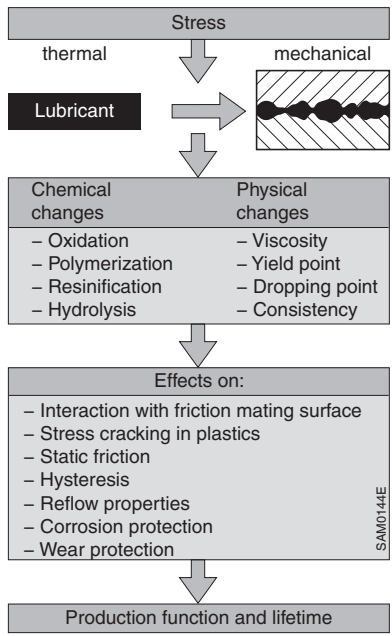


Every chemical change automatically causes a change in physical properties. These include the rheological properties as well as changes in the viscosity-temperature characteristics or the dropping point. A marked lowering of the dropping point would result in the lubricant flowing away from the friction surface, even at moderate heat.

It is particularly important to remember that metals such as iron or copper (or metals containing copper such as bronze or brass) catalyze the oxidation of a lubricant, i.e. oxidation occurs much more quickly than without catalyst contact. Oxidation quickly renders the lubricity of grease insufficient. Often the soap structure decomposes, the grease then becomes oily, flows away from the friction surface, or hardens due to polymerization.

By correctly matching the lubricating grease and tribological system while taking into account the load and interaction, it is possible to enhance the performance potential of products substantially with sliding-contact opposed parts (e.g. transmission, friction or roller bearings, actuator and control systems) [30].

Figure 8: Stress exerted on lubricant and the resulting effects



Lubricants for high voltage requirements

Stronger electric alternating fields than previously occur with increasing operating voltage in alternators and electric motors. This can give rise for example in the ball bearings of motors or alternators to electric discharges between the rolling elements and the raceway. Sparks flash over and can cause tiny areas of the metal to melt. This is called electrical pitting and can cause premature bearing failure. The energy of these discharges will in future be potentially greater if power density and vehicle system voltage increase, e.g. 400 to 1000 V in electric and hybrid vehicles. Without suitable countermeasures there is a high risk of premature bearing damage, initially discernible from increased bearing noise. By reducing the specific resistance of the roller-bearing greases by magnitudes of around 5, for example with ionic liquids, it is possible to reduce dramatically the number and the energy of the discharges.

References

[1] T. Mang, W. Dresel: Lubricants and Lubrication. 2nd Edition, Wiley-VCH Verlag GmbH, 2007.
[2] Thomas Mezger: Das Rheologie-Handbuch. 4th Edition, Vincentz-Network, 2012.
[3] Wilfried J. Bartz: Schmierfette. 2nd Ed., Springer-Verlag, 2016.
[4] DIN 51575: Testing of mineral oils – Determination of sulfated ash (publication 2016).
[5] DIN 51817: Testing of lubricants – Determination of oil separation from greases under static conditions (publication 2014).
[6] DIN EN ISO 2592: Petroleum products – Determination of flash and fire points – Cleveland open cup method (ISO 2592:2000); German version EN ISO 2592:2002.
[7] DIN EN 23015: Petroleum products; determination of cloud point (ISO 3015:1992); German version EN 23015:1994.

[8] DIN 51805: Testing of lubricants; determination of flow pressure of lubricating greases, Kesternich method (publication 1974).
[9] DIN 1342-3: Viscosity – Part 3: Non-newtonian liquids (publication 2003).
[10] DIN ISO 2137: Petroleum products – Lubricating grease and petrolatum – Determination of cone penetration (ISO 2137:1985) (publication 1997).
[11] DIN ISO 3016: Petroleum oils; determination of pour point (publication 1982).
[12] DIN ISO 3448: Industrial liquid lubricants – ISO viscosity classification (ISO 3448:1992) (publication 2010).
[13] DIN 1342-1: Viscosity – Part 1: Rheological concepts (publication 2003).
[14] DIN ISO 2176: Petroleum products – Lubricating grease – Determination of dropping point (ISO 2176:1995) (publication 1997).
[15] DIN EN ISO 3104: Petroleum products – Transparent and opaque liquids – Determination of kinematic viscosity and calculation of dynamic viscosity (ISO 3104:1994 + Cor. 1:1997); German version EN ISO 3104:1996 + AC:1999.
[16] DIN ISO 2909: Petroleum products – Calculation of viscosity index from kinematic viscosity (ISO 2909:2002) (publication 2004).
[17] SAE J300: Engine Oil Viscosity Classification (publication 2015).
[18] SAE J306: Automotive Gear Lubricant Viscosity Classification (publication 2005).
[19] DIN ISO 2137: Petroleum products – Lubricating grease and petrolatum – Determination of cone penetration (ISO 2137:1985) (publication 1997).
[20] ASTM D 2983: Standard Test Method for Low-Temperature Viscosity of Lubricants Measured by Brookfield Viscometer (publication 2015).
[21] ASTM D 445a: Standard Test Method for Kinematic Viscosity of Transparent and Opaque Liquids (and Calculation of Dynamic Viscosity) (publication 2015).
[22] ASTM D 5293: Standard Test Method for Apparent Viscosity of Engine Oils and Base Stocks Between –10 and –35 °C Using Cold-Cranking Simulator (publication 2015).
[23] ASTM D 4684: Standard Test Method for Determination of Yield Stress and Apparent Viscosity of Engine Oils at Low Temperature (publication 2014).
[24] ASTM D 4683: Standard Test Method for Measuring Viscosity of New and Used Engine Oils at High Shear Rate and High Temperature by Tapered Bearing Simulator Viscometer at 150 °C (publication 2013).
[25] CEC L-36-A-90: European standard test for HTHS viscosity, technically identical to ASTM 4741.
[26] ASTM D 4741: Standard Test Method for Measuring Viscosity at High Temperature and High Shear Rate by Tapered-Plug Viscometer (publication 2013).
[27] ASTM D 5481: Standard Test Method for Measuring Apparent Viscosity at High-Temperature and High-Shear Rate by Multicell Capillary Viscometer (publication 2013).
[28] DIN 51502: Designation of lubricants and marking of lubricant containers, equipment and lubricating points (publication 1990).
[29] DIN 51818: Lubricants; consistency classification of lubricating greases; NLGI grades (publication 1981).
[30] P. M. Lugt: Grease Lubrication in Rolling Bearings. 1st Edition, John Wiley & Sons, 2013.

Coolants for internal-combustion engines

Requirements

When a fuel is combusted in the engine, the supplied energy is converted in the proportions of approx. 1/3 into kinetic energy and approx. 2/3 into heat. Roughly 50 % of this heat escapes via the exhaust system, the remainder is utilized or dissipated in the engine cooling circuit via the coolant. The coolant must for this purpose satisfy a series of important requirements:
– optimal heat-transfer properties,
– high heat capacity,
– low evaporative losses,
– good antifreeze properties,
– corrosion, erosion and cavitation protection of all the cooling-circuit materials,
– compatibility of elastomers, plastics, and coatings,
– avoidance of deposits ("fouling") which result in blockages,
– low maintenance requirement, long service life, and easy handling,
– environmental compatibility.

Table 1: Requirements regarding water from the main based on the radiator manufacturer's recommendation

Property	Requirement
Appearance	colorless, clear
Deposit	0 mg/*l*
pH value	6.5 to 8.0
Total hardness	≤ 4.5 mmol/*l*
Chloride	≤ 100 mmol/*l*
Sulfate	≤ 100 mmol/*l*
Hydrogen carbonate	≤ 100 mmol/*l*

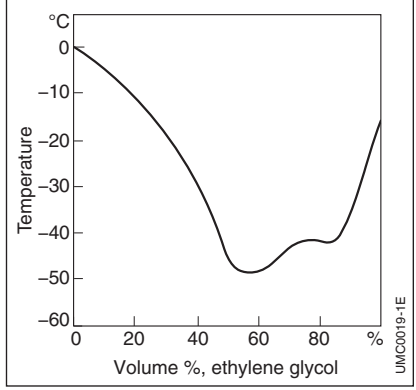
Composition of coolants

Customary coolants for engine cooling circuits consist of a mixture of water with a radiator protection agent approved by the automobile and engine manufacturers. The volume percentage of the radiator protection agent is for the most part between 40 % and 50 %.

Water
Owing to its high specific heat capacity and its correspondingly substantial thermal-absorption capacity, water is a very good coolant. Water from the main is usually used. Since the quality of such water can vary dramatically, the requirements with regard to the water used are specified by most automobile and engine manufacturers, and by radiator manufacturers (Table 1). A higher water ion content results in faster inhibitor consumption and possible after-effects such as, for example, flocculation (coagulation) of low-solubility reaction products, corrosion of circuit components, through to blockage of cooling ducts.

Radiator protection agents
The base material of radiator protection agents is for the most part 1,2-ethane-

Figure 1: Freezing point of coolant as a function of ethylene-glycol content



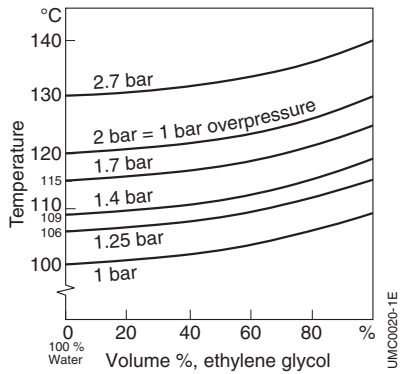
diol (ethylene glycol), or less commonly 1,2-propandeiol (propylene glycol). Both glycols have very good water-soluble properties. When mixed with water from the main, they lower the freezing point and raise the boiling point. The optimum level of these effects is achieved with a volume percentage of approximately 50 % (Figure 1). The boiling point is furthermore dependent on the operating pressure in the cooling system (Figure 2). The operating pressure of a passenger-car cooling circuit is in normal operation typically approximately 1.5 bar overpressure, under full load and in high-performance engines up to 2.5 bar and higher.

In the strict sense, with glycol-water mixtures there is not a defined freezing point, but an ice flaking point at which ice crystals begin to separate out in the fluid. At this temperature, the fluid medium can still be pumped through the cooling circuit.

The heat capacity of glycol-water mixtures is reduced compared with water from the main.

Since water/glycol mixtures demonstrate similar corrosivity to water from the main compared with metals, inhibitor systems are additionally admixed with the radiator protection agents.

Figure 2: Boiling-point curves of water/ethylene-glycol mixtures with different ethylene-glycol content as a function of pressure



Types of radiator protection agents

Radiator protection agents can essentially be classified into three groups, which differ with regard to the inhibitor systems used.

Conventional radiator protection agents containing silicate

Inhibitors and additives

These radiator protection agents contain primarily inorganic inhibitors such as, for example, silicate, nitrite, nitrate or molybdate. Organic inhibitors such as toluene or benzotriazole are also used.

To keep the pH value at the desired alkaline level over the entire period of use, buffer substances such as borate, phosphate, benzoate or imidazole are added to the radiator protection agent.

Radiator protection agents are completed by the addition of cleaning agents (e.g., sulfonates), antifoaming agents, and dyes.

Conventional radiator protection agents containing silicate have proved themselves in practice over decades. Particularly the proportion of silicate has counteracted the dangerous hot corrosion of aluminum engine through the formation of thin, protective coatings.

Service life

However, the drawback of this type of radiator protection agent is that it is not possible to select the concentrations of the inhibitor systems at any level owing to the low solubility of some additives. On the other hand, the inorganic inhibitors are reduced during driving operation; however, the necessary minimum concentrations of inhibitors may not be undershot here. Consequently, the service life of such a radiator protection agent is limited to around three years driving or 100,000 km (for passenger cars). This service life can be significantly reduced for example by the high thermal load of the cooling circuit in modern high-performance engines.

The limited service life of conventional radiator protection agents does not match the wishes of the automobile and engine manufacturers to extend the service intervals of vehicles. The manufacturers of

radiator protection agents are therefore responding with the development of "OAT products".

OAT radiator protection agents
OAT radiator protection agents (Organic Acid Technology) contain a combination of organic inhibitors instead of conventional inorganic inhibitors. Typical combinations are aliphatic mono- and dicarboxylic acids, azelaic acid and aromatic carboxylic acids, sebacic acid and benzotriazole. The combination of an inorganic inhibitor and carboxylic acid is also used. Pure OAT products, however, contain no silicate.

In contrast to conventional radiator protection agents containing silicate, the inhibitors in OAT radiator protection agents decrease must more slowly in driving operation. This enables to service life of the radiator protection agents to be increased and the maintenance requirement to be reduced.

Hybrid radiator protection agents
To combine the good long-term durability of OAT radiator protection agents with the good hot-corrosion protection of silicate inhibitors, hybrid radiator protection agents have been developed which contain a harmonized combination of both inhibitor types. In this way, local hot corrosion can be prevented by the rapid formation of silicate protective coatings, while good long-term corrosion protection, even after the consumption of silicate inhibitors, is provided by the more stable OAT inhibitors.

This on the one hand satisfies the demand for low-maintenance coolants, but on the other hand also takes into account the high thermal loads of modern engine technologies. Therefore these hybrid radiator protection agents, sometimes also called Si-OAT radiator protection agents, are today frequently used by major vehicle manufacturers for the initial fill.

Brake fluids

Requirements

Passenger cars and light commercial vehicles are equipped with hydraulic brake systems. A hydraulic medium transmits and converts the braking force from the brake pedal in the brake master cylinder into hydraulic pressure, which is routed via brake lines to the wheel brake and there converted back into braking force. Vehicle-dynamics control systems such as the antilock braking system or the electronic stability program can additionally modulate the hydraulic pressures at each wheel to prevent the wheels from locking through pressure reductions or to stabilize the handling through additional pressure build-ups. This medium is therefore subject to the following requirements.

Low compressibility
The basic requirement with regard to the hydraulic medium brake fluid for effective power transmission is very low compressibility. The liquid state of aggregation must be maintained in the entire operating range for this purpose.

Water-absorption capacity
During driving water can penetrate into the brake system at the wheels, particularly through the flexible brake hoses usually made from rubber material. If this water is not bound by the brake fluid, compressible water-vapor bubbles can form at temperatures above 100 °C. Consequently, no further pressure can be generated when the brake pedal is pressed because initially the compressible water vapor is compressed. This failure of the brakes is called “vapor lock”. Consequently, the market has seen the widespread acceptance of brake fluids which are completely mixable with water and exclude the presence of unbound, free water. Mineral-oil-based brake fluids also commonly used in the past have in practical terms been completely superseded by chemically glycol-based brake fluids (market share is approximately 99 %).

Standards for brake fluids
Brake fluids must comply with stringent requirements to ensure reliable brake-system operation. Their properties are defined in various standard requirements which are very similar in terms of content (SAE J1703 [1], SAE J1704 [2], FMVSS 116 [3], ISO 4925 [4], JIS K2233 [5]). Exclusively for the vehicle operating medium brake fluid the performance requirements of FMVSS 116 (Federal Motor Vehicle Safety Standard) became mandatory in the USA and often serve as a universal reference. In this standard the US Department of Transportation (DOT) – comparable with other standards – has defined specific ratings for salient properties (Table 1).

Table 1: Classification of brake fluids with different chemical bases

Chemical base / Classification	Glycol ether				Silicone	Mineral oil
Suggested color	Colorless to amber				Purple	Green
FMVSS 116	DOT 3	DOT 4		DOT 5.1	DOT 5	
ISO 4925 class	3	4	6	5-1		ISO 7308
SAE	J 1703	J1704 Standard	J1704 Low viscosity		J 1705	
JIS K2233 class	3	4	6	5		
Dry boiling point [°C] Wet boiling point [°C]	>205 >140	>230 >155	>250 >165	>260 >180	>260 >180	>235
Viscosity at –40 °C [mm²/s]	<1500	<1800 (DOT) <1500 (SAE, ISO, JIS)	<750	<900	<900	<2000

Characteristics

Boiling point
During braking the temperature at the wheel brake cylinders increases, in extreme cases (e.g., frequent braking on long downhill gradients) even into ranges around 200 °C. Vapor bubbles can form at temperatures above the operating medium's instantaneous boiling point. The brake fluid transforms at least partie into a vapor state that is too compressible for reliable braking operation. It is then no longer possible to operate the brakes as is the case with the presence of water-vapor bubbles; "vapor lock" brake failure occurs.

Dry boiling point
In their new state (e.g. in sealed containers) brake fluids typically contain no or very little water, i.e. the water content is typically < 0.2 %. Fluids, particularly in steel containers, sometimes reach values of < 0.03 %. The equilibrium boiling point in this new state is called the dry boiling point (ERBP: Equilibrium Reflux Boiling Point).

The boiling point of the hygroscopic glycol-based brake fluid decreases significantly as a result of water absorption. Consequently, the risk of a "vapor lock" increases accordingly at high wheel-brake temperatures.

Depending on operation and the climate region, typically a water-content increase of approximately 1 to 1.5 % per year is assumed.

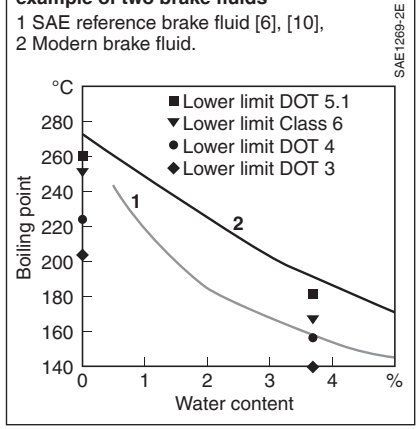
Wet boiling point
The equilibrium boiling point of the brake fluid after a precisely specified procedure of water absorption is defined as the wet boiling point (wERBP: wet Equilibrium Reflux Boiling Point) and is intended to depict the boiling point of a brake fluid used for two to three years.

The brake fluid to be tested is removed after roughly one day from a very humid climatic chamber to measure the boiling point when a reference brake fluid has reached a 3.7 % water content. The boiling point of the operating medium continues to drop if further water is absorbed. Thus the brake-failure risk increases at high wheel-brake temperatures. This is the main reason why the brake fluid typically should be replaced every two years. Figure 1 shows by way of example the drops in boiling point that result from water absorption in two brake fluids.

A significant development trend on the part of the brake-fluid manufacturers in previous years was to increase the boiling point with increased water content.

Viscosity
The viscosity describes the resistance to flow of the operating medium. As the viscosity increases, so too the pressure drops increase as the medium flows through lines and throttle points, e.g., in valves and pumps of the vehicle-dynamics control systems. This reduces, for example, the dynamics of the pressure build-up, which can reduce the effectiveness of the electronic stability program. Therefore, generally speaking, low cold viscosities and low dependence of viscosity on temperature are desirable. This is specified by upper limits for the viscosity at very low temperatures at –40 °C (cold viscosity). It should be as low as possible in order to facilitate high dynamics on the part of the pressure build-up. Reducing the viscosity at very low temperatures was a further development trend on the part of the brake-fluid manufacturers in past years. Figure 2 shows by way of

Figure 1: Reduction of the boiling point with increasing water content from the example of two brake fluids
1 SAE reference brake fluid [6], [10],
2 Modern brake fluid.



example the dependence of viscosity on temperature for two brake fluids. This has a clearly non-linear characteristic particularly at low temperature. As a general "rule of thumb": a 1 K reduction in temperature produces approximately a 10 % increase in viscosity. Likewise, approximately a 10 % increase in viscosity results at low temperature additionally from a 1 % increase in water content. This effect is another reason why the brake fluid should be changed.

Corrosion protection
FMVSS 116 (and also all the other standards) stipulates that brake fluids shall exercise no corrosive effect on those metals generally employed in brake systems. Appropriate additives guarantee the required corrosion protection, which is checked beyond required upper limits in high-temperature storage tests of metal plates in dry and in wet brake fluids.

Elastomer swelling
Elastomers are used as seals and gaskets in the brake system, for example in the brake master cylinder, in vehicle-dynamics control systems, and in hoses. Although a small amount of elastomer swelling is desirable and taken into account in the design of the sealing rings, it should not exceed 10 % under any circumstances; otherwise, it has a negative effect on the strength of these components.

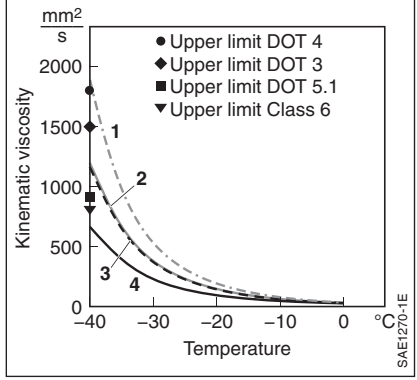
Typically, EPDM is used as elastomer material for glycol-based (and also for rarely used silicone-oil-based) brake fluids. Mineral-oil-based brake fluids are not compatible with EPDM and required different elastomer materials (e.g., FKM).

Even minute levels of mineral-oil contamination (e.g., hydraulic fluids, mineral-oil-based brake fluid) of glycol-based brake fluid can destroy EPDM elastomers (e.g., sealing elements). This can ultimately lead to brake failure. To avoid such contamination, high standards of cleanliness must be maintained for the brake fluids used during service filling [10].

Chemical composition

Glycol-ether fluids
Most brake fluids are based on glycol-ether compounds. These generally consist of monoethers of low polyethylene glycols. The chemical formulation of the ISO and at the same time the SAE reference brake fluids is specified in detail in ISO 4926 [6]. Glycol components can be used to produce a brake fluid which conforms to DOT 3 requirements (Table 1), the downside is that their boiling point drops relatively quickly with water absorption. DOT 4 and DOT 5.1 brake fluids are characterized by the fact that glycol-ether borates are also used. This enables the water to be chemically bound. In this way, the boiling point drops much more slowly compared with DOT 3 fluids. The wet boiling point rises significantly (Figure 1). Thus, the risk of a "vapor lock" decreases significantly when the wheel brake is subject to increased water content and high temperatures.

Figure 2: Dependence of viscosity on temperature and water content from the example of two brake fluids
1 SAE reference brake fluid with 5 % water,
2 SAE reference brake fluid without water [6], [10],
3 Modern brake fluid with 5 % water,
4 Modern brake fluid without water.



Future development trends
ISO 4925 defines a further quality class, namely “Class 6”. This quality is better than that of DOT 4 and is characterized by its particularly low cold viscosity (Table 1). It constitutes the first combination of the two development trends of increased wet boiling point and reduced cold viscosity.

This trend towards higher wet boiling points and low cold viscosities will continue and combine. Increasingly “modern” brake fluids are being developed and offered.

In addition, the functionality of the vehicle-dynamics control systems – e.g., vehicle-speed controller including brake intervention as ACC (Adaptive Cruise Control) and AEB (Automated Emergency Braking) – is constantly increasing. The road towards (semi-)automated driving is clear. The frequency of use of the vehicle-dynamics control systems thereby increases. While a pump element in the ABS (antilock braking system) was only rarely active, its duration of use and hence also its operating load increase significantly by way of TCS (traction control system) and electronic stability program with comprehensive additional functions right up to automated driving.

This also renders necessary in future in the brake-fluid standards requirements, e.g., with regard to the lubricity of the brake fluid, in order to reduce wear. The same also applies to avoid noise when operating the brake master cylinder and clutch, which in the meantime is increasingly operated in manual transmissions with brake fluid. Working groups, particularly within the SAE and ISO, and the pfp TriNoWe (public funded project Tribology in Standards Worldwide; German: Tribologie in Normen Weltweit), are working on further developments of the standards.

Other types of brake fluids
Silicone-oil fluids (SAE J1705, [7])
Silicone-oil-based brake fluids are now only used in a few special applications, e.g. in motor sport, in which very short maintenance intervals are possible.

The disadvantage of these products is – aside from the high price – the higher compressibility. Furthermore, they are – like mineral oils – not hygroscopic and therefore not able to absorb water that penetrates into the brake system. For this reason, they are also called “Low Water Tolerant” brake fluids.

Mineral-oil fluids (ISO 7308, [8])
Additives for improving the fluid properties (e.g., lubrication) are known from mineral-oil technology. These were used previously for example in vehicles with central hydraulics, but are now no longer significant because they – like silicone-based fluids – cannot bind water.

Mineral-oil fluids should never be added to brake systems which are designed for glycol fluids (or vice versa), as this would destroy the elastomer seals. For this reason, the FMVSS suggests a color-coding system for the various brake fluids (see Table 1).

References
[1] SAE J1703: Motor Vehicle Brake Fluid.
[2] SAE J1704: Motor Vehicle Brake Fluid Based Upon Glycols, Glycol Ethers and the Corresponding Borates.
[3] FMVSS 116: Federal Motor Vehicle Standard No. 116: Motor Vehicle Brake Fluids.
[4] ISO 4925: Road vehicles – Specification of non-petroleum-base brake fluids for hydraulic systems.
[5] JIS K2233: Non-petroleum base motor vehicle brake fluids.
[6] ISO 4926: Road vehicles – Hydraulic braking systems – glycol-based reference brake fluids.
[7] SAE J1705: Low Water Tolerant Brake Fluids.
[8] ISO 7308: Road vehicles – Petroleum-based brake fluid for stored-energy hydraulic brakes.
[9] SAE J1709: (Historical) European Brake Fluid Technology.
[10] SAE J1706: Production, Handling and Dispensing of SAE J1703 Motor Vehicle Brake Fluids and J1704 Borate Ester Based Brake Fluids.

Fuels

Overview

Composition

Fuels consist of a multitude of hydrocarbons, primarily of various paraffins and aromatic compounds.

Paraffins

Paraffins are linear, isoparaffins branched and cycloparaffins closed-chain hydrocarbons in which the atoms are linked to each other by single bonds. Molecules with single bonds are chemically less reactive.

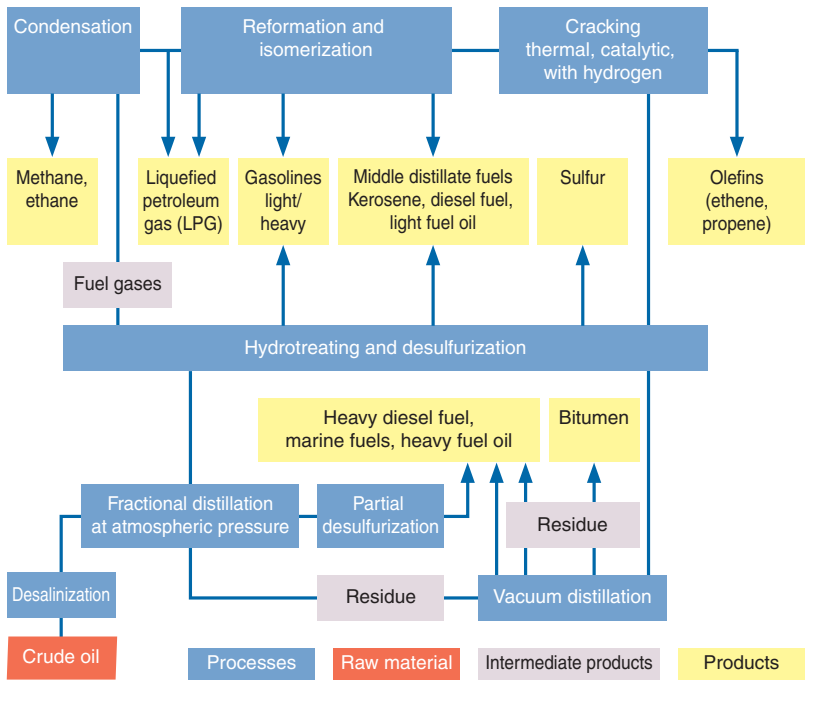
Aliphatic compounds

Paraffins belong to the class of aliphatic compounds. Aliphatic compounds cover all linear, branched and closed-chain hydrocarbons whose molecules have one or more single bonds, one or more isolated multiple bonds or a combination of both types of bond. Molecules which exhibit an aromatic system are not classed as aliphatic compounds.

Aromatic compounds

Aromatic compounds are characterized by the property that the molecules have at least one ring, in which electrons are distributed around the entire ring. Benzene is a typical example. Aromatic compounds are likewise less active, because the elec-

Figure 1: Oil-refining processes in a refinery



trons are not located as a double bond between two carbon atoms.

Manufacture

Gasoline and diesel fuel are produced from crude oil. First the crude oil is desalted and distilled, where the hydrocarbons are separated, depending on the molecule size, into different fractions (Figure 1).

Distillation under atmospheric pressure

The process is begun with a distillation of the crude oil under atmospheric pressure, in which different substance streams of predominantly volatile components are produced. Depending on their composition, these substance streams form the basis for liquefied petroleum gas (propane and butane), gasoline, kerosene, and for the more volatile part of diesel fuel and fuel oil.

Distillation under vacuum

The remaining residue from the atmospheric distillation is further separated under vacuum. These fractions form the basis for the heavier components in diesel fuel and fuel oil and for light and heavy marine fuels. During vacuum distillation too a residue is left, which is processed into heavy fuel oil and bitumen.

Further processing

The ratio and the quality of the product fractions obtained by distillation do not yet meet market requirements such that further processing in complex chemical processes is required in the refinery.

Aside from the widely known desulfurization, larger molecules can be cracked purely thermally or in the presence of catalysts (thermal cracking, catalytic cracking). If this process is performed in the presence of hydrogen, it is known as hydrotreating or hydrocracking. Additional shorter-chain components can be produced from heavy oil for diesel fuel by hydrocracking.

The structure of molecules is modified in other refinery processes. During reforming aromatic compounds are produced from saturated aliphatic molecules, where hydrogen that is needed for hydrotreating or hydrocracking and desulfurization is created. Linear molecules can be converted into branched molecules by isomerization processes. These increase the octane number of gasoline and the low-temperature resistance of diesel fuel.

The interaction of the complex handling processes in a refinery are shown in greatly simplified schematic form in Figure 1.

Addition of biofuels and additives

Today proportions of biofuels are added to many fossil fuels worldwide. Certain properties of biofuels can be specifically used to satisfy the fuel-grade requirements. For example, bioethanol increases the octane number of gasoline and adding biodiesel to diesel fuel ensures sufficient lubricity. Additives are often also used to further improve product quality. Additives are active substances which are added to specifically adjust certain fuel properties.

Characteristics

Boiling range and ignition temperature
The boiling range of gasoline is between 30 °C and 210 °C, and that of diesel fuel between 160 °C and 370 °C. Diesel fuel ignites on average at approximately 350 °C (lower limit 220 °C), which is very early in comparison with gasoline (on average 500 °C).

Net calorific value
Normally, the net calorific value H_i (former term: lower calorific value) is specified for the energy content of fuels; it corresponds to the usable heat quantity released during full combustion.

The gross calorific value H_s (former term: upper calorific value), on the other hand, specifies the total reaction heat released and therefore comprises, as well as the mechanically usable heat, the latent heat created in the water vapor. However, this component is not used in the vehicle.

The net calorific value H_i of diesel fuels is in the range of 42.9 through 43.1 MJ/kg and is slightly higher than the calorific value of gasoline (40.1 through 41.9 MJ/kg).

Fuels or fuel constituents containing oxygen (oxygenates) such as alcohols (H_i of ethanol: 26.8 MJ/kg) and fatty acid methyl esters (H_i of biodiesel: 36.5 MJ/kg) have a lower calorific value than fuels which only contain hydrocarbons. The net calorific value is reduced because bonded oxygen does not contribute to the release of energy during combustion. Engine performance comparable with conventional fuels can therefore only be achieved with fuels containing oxygen by increased fuel consumption. Subsequently increasing the ethanol content in gasoline above 10 % (V/V) would in many existing vehicles result in a noticeable reduction of performance and in significantly higher fuel consumption.

Calorific value of air/fuel mixture
This is the calorific value of the combustible air/fuel mixture. It is dependent on the air/fuel ratio and determines the engine's power output. With a stoichiometric air/fuel ratio, this is roughly 3.5 - 3.7 MJ/m^3 for all liquid fuels and liquefied petroleum gases.

Sulfur content
In the interests of reducing sulfur-dioxide emissions (SO_2) and protecting the catalytic converters for exhaust-gas treatment, the sulfur content of gasolines and diesel fuels was limited for vehicle applications on a Europe-wide basis to 10 mg/kg as from 2009. Fuels which adhere to this limit value are known as "sulfur-free fuels". In this way, the final stage of the desulfurization of fuels is achieved.

Since 2017 in the US too the average sulfur content of gasoline which is supplied commercially to end consumers has been no more than 10 mg/kg, with individual exceptions being permitted, e.g., up to 80 mg/kg for refineries and up to 95 mg/kg in downstream areas. All sulfur-free diesel fuels (Ultra Low Sulfur Diesel, ULSD) for on- and off-road vehicles may not contain more than 15 mg/kg sulfur in the US.

In China the sulfur content of gasolines and diesel fuels may not exceed 10 mg/kg.

With the next stage of emissions legislation for on-road vehicles in India the fuels containing up to 50 mg/kg sulfur still available on the market have since 2020 been gradually replaced by sulfur-free products.

Generally, a marked increase in the desulfurization of fuels can be observed worldwide.

Gasoline

Fuel grades

Super fuels in Germany

In Germany two Super fuels with 95 octane are available; these differ in their ethanol content and may contain a maximum 5 % (V/V, % by volume) ethanol (Super) or 10 % (V/V) ethanol (Super E10). A Super Plus fuel with 98 octane is also available. Individual providers markets fuels with an octane number of at least 100 (V-Power Racing 100, Super 100plus, Ultimate 102), which in comparison with the Super fuels with 95 octane are changed in terms of basic quality and additives.

Fuels in the USA

Three different types of fuel are sold in the USA: Regular (92 octane), Premium (94 octane), and Premium Plus (98 octane). Fuels sold in the USA usually contain 10 % (V/V) ethanol. The addition of components containing oxygen increases the octane number and satisfies the requirements of modern, higher-compression engines with regard to improved knock resistance.

Reformulated gasoline is the term used to describe gasoline which, through its altered composition, generates fewer evaporative and exhaust emissions than conventional gasoline. The demands placed on reformulated gasoline are laid down in the USA in federal law (Clean Air Act). In comparison with standard quality, this legislation prescribes, for example, lower limits for vapor pressure, benzene and overall aromatic content, and final boiling point. It also prescribes the use of additives to keep the intake system free of contamination and deposits (detergents).

Fuel standards

European Standard EN 228 [1] defines the requirements for unleaded fuel for use in gasoline engines (Table 1). Further country-specific characteristics are set out in the national appendices to this standard. Leaded gasolines are no longer permitted in virtually all markets worldwide and by 2020 were only available in a few countries, e.g., Algeria and Iraq.

The US specifications defining fuels for gasoline engines are contained in ASTM D4814 (American Society for Testing and Materials) [2].

Most fuels for gasoline engines which are sold today contain components which in turn contain oxygen (oxygenates). Ethanol has gained in importance in Europe as well, since the “EU Biofuels Directive” stipulates minimum content figures for renewable fuels. But also those ethers

Table 1: DIN EN 228 (August 2017): Selected requirements of gasolines

Requirements	Unit	Characteristic	
Quality		Super E10	Super
Knock resistance: RON/MON for Super, min.	–	95/85[1]	
Density (at 15 °C), min./max.	kg/m^3	720/775	
Sulfur, max.	mg/kg	10	
Oxygen content, max.	% (m/m)	3.7	2.7
Ethanol content, max.	% (V/V)	10.0	5.0
Benzene, max.	% (V/V)	1	
Lead, max.	mg/*l*	5	
Volatility			
Vapor pressure, summer, min./max.	kPa	45/60	
Vapor pressure, winter, min./max.[2]	kPa	60/90	
Evaporated volume at 70 °C, summer, min./max.	% (V/V)	22/50	20/48
Evaporated volume at 70 °C, winter, min./max.	% (V/V)	24/52	22/50
Evaporated volume at 100 °C, min./max.	% (V/V)	46/72	46/71
Evaporated volume at 150 °C, min./max.	% (V/V)	75/–	
Final boiling point, max.	°C	210	
VLI[3] transitional period[4], max.[2]		1 164	1 150

[1] Applicable to Super Plus: RON/MON, min. 98/88; ethanol content, max. 10.0 % (V/V).
[2] National values for Germany. [3] VLI Vapor-lock index. [4] Spring and fall.

which can be produced from methanol or ethanol MTBE (methyl tertiary butyl ether) and ETBE (ethyl tertiary butyl ether) are used, of which, depending on the quality of the basic fuel, up to 22 % (V/V) may be added in Europe. A mixture of alcohols and ethers is also permitted. In Europe the overall quantity of oxygenates is limited by the stipulation of a maximum permitted oxygen content. Thus, E10 in Europe may exhibit a maximum oxygen content of 3.7 % (m/m, mass percent), which is already exhausted with the addition of 10 % (V/V) ethanol. When ether is added, the maximum permitted volume percentage of ethanol must be reduced accordingly. Worldwide many countries have defined minimum quotas for biogenic components in gasolines, which are achieved for the most part with bioethanol.

Adding alcohols increases the volatility at higher temperatures. Furthermore, alcohols can damage materials in the fuel system, lead for example to elastomer swelling, and trigger alcoholate corrosion on aluminum parts. Depending on the alcohol content and the temperature, demixing may result in response to only small amounts of water. During phase separation a new aqueous phase is formed with part of the alcohol from the fuel. The use of fuels containing alcohol therefore require additional measures on the vehicle and in the filling-station network.

The ethers do not encounter the problem of demixing. The ethers, which have a lower vapor pressure, a higher calorific value and a higher octane number than ethanol, are chemically stable components with good material compatibility. They therefore demonstrate advantages, both from an engine-engineering and a logistical standpoint. For reasons of sustainability, because of the CO_2 savings to be attained and due to the laying down of quotas for biogenic fuels, ETBE is used in the main.

The ethanol content in the European gasoline standard EN 228 was for a long time limited to 5 % (V/V) (E5). In 2013 EN 228 was in the first instance supplemented by a specification for 10 % (V/V) ethanol (E10). At present in the European market not all vehicles are yet equipped with materials to allow operation with E10. As the second quality, therefore, an exemption grade with a max. ethanol content of 5 % (V/V) is retained. E10 is conceived as the main grade in the gasoline market. E10 has failed to find favor with consumers in some of the European countries in which it has been introduced. E10 is primarily favored In Bulgaria, Romania, Slovakia, the Netherlands, Belgium, Luxembourg, and Finland, but in Germany is well below expectations with a market share of 13 % (as at April 2020).

In the USA most gasolines contain ethanol at 10 % (V/V) (E10), for which a vapor pressure higher by approximately 7 kPa is permitted according to the American standard ASTM D4814. If certain preconditions are satisfied, a vapor pressure that is up to 8 kPA higher is also according to EN 228, depending on the ethanol content.

In the USA the amount of ethanol that may be added to gasoline is limited by law to 10 % (V/V). Because there are increasing demands for this limitation to be dropped, the upper limit for ethanol has in ASTM D4814 been increased as a precaution to 15 % (V/V). The vapor-pressure increase of 7 kPa, which until now has been confined to E10, will also apply in future to E15.

Characteristics

Density

European standard EN 228 fixes the density of gasolines at 720 – 775 kg/m^3.

Octane number

The octane number defines the gasoline's antiknock quality (resistance to pre-ignition). The higher the octane number, the greater the resistance to engine knock. Iso-octane (trimethyl pentane), which is extremely knock-resistant, is assigned the octane number 100, while n-heptane, which is extremely knock-susceptible, is assigned the number 0.

The mixture of iso-octane and n-heptane, which exhibits the same knock resistance as the fuel to be tested, is determined in a standardized test engine. The proportion in % by volume of iso-octane in the mixture of iso-octane and n-heptane corresponds to the fuel's octane number.

Iso-octane has a high octane number, but there are components which are even more knock-resistant, for example toluene (RON = 124) or ethanol (RON = 130). The definition and verification of octane numbers above 100 is based on primary reference standards from iso-octane with different proportions of tetraethyl lead. To determine the amount of tetraethyl lead required for a specific octane number above 100, it was necessary to carry out comprehensive examinations of the knock characteristics of mixtures which consisted of benzene (RON = 99), varying proportions of iso-octane and heptane as well as tetraethyl lead.

Research and motor octane numbers
RON ((research octane number) is the octane number determined in testing using the Research Method [3]. It serves as the essential index of acceleration knock. The octane number determined in testing using the Motor Method [4] is the MON (motor octane number). The MON basically provides an indication of the tendency to knock at high speeds.

The Motor Method differs from the Research Method by using preheated mixtures, higher engine speeds and variable ignition timing. This places more stringent thermal demands on the fuel under examination. MON figures are lower than those for RON.

Increasing the antiknock quality
Normal (untreated) straight-run gasoline displays a low antiknock quality. Only by mixing such gasoline with different knock-resistant refinery components (reformed components, isomerisates) is it possible to produce fuels with high octane numbers which are suitable for modern, high-compression engines. Even components containing oxygen such as alcohols and ether increase the antiknock quality. They are therefore specifically added for this purpose.

Additives containing metal to increase the octane number, e.g., MMT (methylcyclopentadienyl manganese tricarbonyl), form ashes during combustion. The addition of MMT is therefore excluded in EN 228 by the fixing of a limit value for manganese in the trace range.

In view of the optimization of downsizing and turbocharging of direct-injection gasoline engines there are increasingly indications that high characteristic RON and MON values alone cannot guarantee actual sufficient knock resistance of fuels in the lower torque range. The pre-ignition behavior can also vary greatly for a similar RON such that an additional characteristic, the compression pre-ignition number (CPI), is suggested to characterize the tendency to pre-ignition [5].

Fuel volatility
The volatility of gasoline has both upper and lower limits. The lower limitation is chosen so that the fuel contains sufficiently volatile components to ensure a reliable cold start. However, the fuel's volatility may not be so high that at higher temperatures the fuel supply is interrupted by the formation of gas bubbles (vapor lock) and as a result difficulties when hot starting or problems when driving are encountered. Yet another factor is environmental protection, which demands that evaporative losses be kept low.

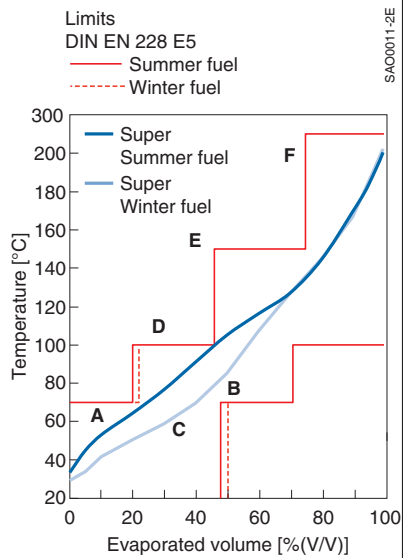
Fuel volatility is defined by different characteristic quantities. EN 228 defines for E5 and E10 ten different volatility classes distinguished by various levels of boiling curve, vapor pressure and VLI (vapor-lock index). To meet special requirements stemming from variations in climatic conditions, European countries can incorporate specific individual classes into their own national appendices in the standard. Different values are laid down for summer and winter.

A similar provision is included in the American gasoline standard ASTM D4814. Comprehensive tables describe the scope of application of the volatility classes, depending on the calender month and state.

Boiling curve
In order to assess the fuel in vehicle operation, it is necessary to view the individual areas of the boiling curve separately. EN 228 therefore contains limit values laid down for the volume of fuel that vaporizes at 70 °C, at 100 °C and at 150 °C. Up to 70 °C an minimum volume of fuel must be vaporized in order to ensure easy starting of the engine when cold (Figure 2, A). Suf-

ficient volatility in the lower temperature range is particularly important for vehicles in which the fuel preparation is still performed by means of a carburetor (Figure 2, B). However, the volume of fuel that vaporizes must not be too great either, otherwise vapor bubbles may be formed when the engine is hot and the evaporative emissions are too high (Figure 2, C). The volume of fuel that vaporizes at 100 °C determines, aside from the engine's warm-up characteristics, primarily the readiness for operation and accelerating performance of the engine at normal operating temperature, otherwise the non-volatile volume in the intake manifold condenses (Figure 2, D). The volume of fuel that vaporizes at up to 150 °C should be high enough to minimize dilution of the engine oil. In particular when the engine is cold, the non-volatile gasoline components find it difficult to vaporize and can pass from the combustion chamber via the cylinder walls into the engine oil (Figure 2, E). Too many non-volatile components in the uppermost range of the boiling curve can cause formation of residues, coking, and sooting (Figure 2, F).

Figure 2: Boiling curve of gasoline according to ISO 3405 [6]
Explanations for A to F see text.



Vapor pressure

Fuel vapor pressure as measured at 37.8 °C (100 °F) in accordance with EN 13016-1 [7] is primarily a characteristic quantity with which the safety requirements in the vehicle's tank are established. Vapor pressure has upper and lower limits in all specifications. In Germany, for example, it is max. 60 kPa in summer and max. 90 kPa in winter.

In order to configure a fuel-injection system, it is also important to know the vapor pressure at higher temperatures (80...120 °C) because a rise in the vapor pressure due to the admixture of alcohol becomes apparent only at higher temperatures. When comparing the vapor-pressure curves of gasolines with different alcohol contents it becomes clear that the amount of alcohol at moderate temperatures (37.8 °C) barely affects the vapor pressure, while fuels with 10 % (V/V) ethanol at 130 °C exhibit a vapor pressure that is higher by more than 200 kPa (Figure 3). Fuels containing 20 to 30 % (V/V) ethanol achieve the highest vapor pressure (up to 800 kPa at 130 °C). Higher ethanol contents, however, lead to a reduction of the vapor pressure again because the intermolecular interactions increase under the polar ethanol molecules and these bind in a more pronounced manner. Methanol fuels develop at high temperatures even higher vapor pressures (vapor pressure at 130 °C of E15: 768 kPa; of M15: 1070 kPa).

If the fuel vapor pressure rises above the fuel-injection system pressure, for example during vehicle operation due to the effect of the engine temperature, this may result in malfunctions caused by vapor-bubble formation. Due to the increased volatility of alcohol fuels, especially for operation with M15, the validation range for system functions and vehicle applications must be checked and if necessary adapted.

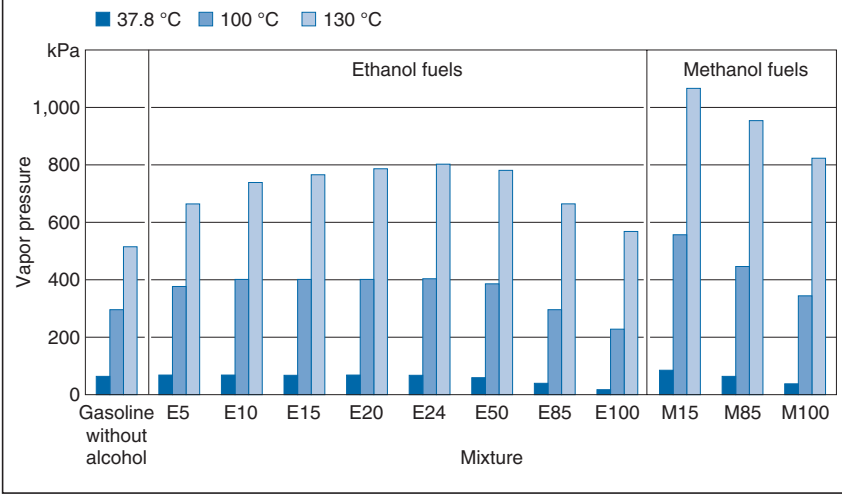
Vapor/liquid ratio
The vapor/liquid ratio is a measure of a fuel's tendency to form bubbles. It refers to the volume of vapor generated by a specific quantity of fuel at a defined back pressure and a defined temperature. A drop in pressure (e.g., when driving over a mountain pass), or an increase in temperature, will raise the vapor/liquid ratio and with it the possibility of operating problems. ASTM D4814 lays down, for example for each volatility class, a temperature at which a vapor/liquid ratio of 20 must not be exceeded.

Vapor-lock index
The vapor-lock index (VLI) is the mathematically calculated sum total of ten times the vapor pressure (in kPa at 37.8 °C) and seven times the volume percentage of fuel that vaporizes at up to 70 °C. It is possible with this additional limit value to restrict the volatility of the fuel further with the result that the two maximum values for vapor pressure and boiling data cannot be achieved in the course of its production.

Additives
Additives can be added to improve fuel quality in order to counteract deteriorations in engine performance and in the exhaust-gas composition during vehicle operation. The packages generally used combine individual components with various attributes.

Care and precision are required both when testing additives and in determining their optimal compositions and concentrations. Undesirable side-effects must be avoided. Basic additives are added at the refinery to protect the plants and to ensure a minimum fuel quality. To further improve quality, brand-specific multifunction additives can be added at the refinery's filling stations when the road tankers are filled (end-point dosing).

Figure 3: Vapor pressure of gasoline as a function of temperature depending on the ethanol and methanol content (DIN EN 13016)



Detergents
The setting and preparation of the mixture composition of air and fuel in the combustion chamber is pivotal for trouble-free driving and minimization of pollutants in the exhaust gas. Keeping the entire intake system (fuel injectors, intake valves) clean is an important prerequisite for meeting the optimized conditions in the new state. Effective cleaning additives (detergents) help to keep surfaces clean and remove existing coatings.

Corrosion inhibitors
The ingress of water/moisture may lead to corrosion in fuel-system components. Corrosion can be effectively eliminated by the addition of corrosion inhibitors, which form a thin protective film on the metal surface.

Aging stabilizers
Oxidation inhibitors (antioxidants) are added to fuels to improve their stability during storage. They prevent rapid oxidation caused by oxygen in the air.

Blend components for fossil gasolines
Bioethanol
Production from sugar and starch
Bioethanol can be obtained from products containing sugar and starch and is the most widely produced biofuel worldwide. Plants containing sugar (sugar cane, sugar beet) are fermented with yeast, the sugar fermenting to form ethanol in the process.

When bioethanol is obtained from starch grains such as corn, wheat or rye are pretreated with enzymes in order to partially split the long-chain starch molecules. A split into dextrose molecules occurs during the subsequent likewise enzymatically performed saccharification. Bioethanol is created in a further process step by means of fermentation with yeast.

Production of lignocellulose
Enzymes can also be used to produce bioethanol from lignocellulose. Lignocellulose forms the structural framework of the plant cell wall and contains the chief constituents lignin, hemicellulose and cellulose. The advantage of this process is that the entire plant can be used and not just that part containing sugar or starch. Because of this new approach, this product is also referred to as 2nd-generation bioethanol. However, the enzymes can split cellulose, but not lignin. The enzymes developed to date are still highly sensitive with regard to the usable biomass such that this process is slow in acquiring economic significance.

Use of bioethanol
Bioethanol is, on the basis of its properties, highly suitable for admixing in gasolines, particularly in order to increase the octane number of pure gasoline. That is why virtually all gasoline standards permit the addition of ethanol as a blend component. Even the biofuel policy of the European Union leads one to expect that the market penetration and the proportion of bioethanol in gasolines will continue to rise if it is guaranteed that bioethanol is created on a sustained basis and not in competition with foods.

Bioethanol can also be used as a pure fuel in gasoline engines in flexible-fuel vehicles (FFVs). These vehicles can run both on gasoline and on pure ethanol as well as any mixture of gasoline and ethanol. A gasoline with 85 % (V/V) ethanol is offered in some European countries, for example Germany, France, and Sweden. On account of the cold-start problems at low temperatures, the maximum ethanol concentration is sometimes reduced depending on the local climate. Fuel samples exhibit typical ethanol contents of 75 to 85 % (E85) in summer and 65 to 85 % in winter. The quality of E85 is defined for Europe in the fuel specification EN 15293 [8] and in the USA in ASTM D5798 [9]. In Brazil gasolines are basically only sold as ethanol fuels, which are also called gasohol and contain 22 to 26 % (V/V) ethanol. Unblended ethanol, E100, is also available at many Brazilian filling stations, but this contains 7 % (m/m) water.

Methanol
Manufacture
Methanol is primarily produced not by regenerative means from fossil energy sources such as coal and natural gas, and therefore makes no contribution to reducing CO_2 emissions. Nevertheless, China for strategic reasons relies on methanol fuels because they enable part of its high fuel demand to be covered locally from coal. Here, M 15 appears to represent an upper limit for use in conventional gasoline engines. In China, an M85 analogous to E85 is additionally being discussed for flexible-fuel vehicles.

Use of methanol
With the same alcohol content, methanol fuels have a significantly greater corrosive effect on ferrous alloys than ethanol fuels. Even demixing occurs much more quickly when the fuel is fouled by water. Because of the negative experiences of methanol fuels in Germany during the oil crisis in 1973 and also due to its toxicity, the use of methanol as a blend component has again been abandoned in Germany. From a worldwide point of view, only occasional methanol admixtures are being conducted at present, and then for the most part – apart from China – with a content of well below 5 % (M5).

Diesel fuel

Diesel-fuel grades
As well as petroleum-based conventional diesel fuel, purely paraffinic fuels, particularly from renewable sources, and biodiesel have also become important.

Fuel standards
The standard that lays down the requirements for conventional diesel fuels in Europe is EN 590 [10]. The most important characteristics are set out in Table 2. Even the premium qualities additionally sold at some filling stations (e.g., Super diesel, Ultimate, V-Power diesel) satisfy this standard. Standard and premium diesel fuels differ in the basic quality, in the addition of paraffinic fuel components, and in the additives.

Up to 7 % (V/V) biodiesel (FAME) can be admixed with diesel fuel in accordance with EN 590, where the quality of the biodiesel is stipulated by standard EN 14214 [11]. The admixture of biodiesel improves lubricity, but also reduces oxidation stability. For the purpose of safeguarding sufficient oxidation stability, EN 590 was supplemented in 2009 to include the parameter induction time and a minimum requirement established. The induction time is a measure of the fuel's aging reserve.

The US standard for diesel fuels, ASTM D975 [12], specifies a smaller number of quality criteria and defines less stringent limits. It permits the admixture of max. 5 % (V/V) biodiesel, which must satisfy the requirements of standard ASTM D6751 [13].

Characteristics
Cetane number
The cetane number (CN) expresses the ignition quality of the diesel fuel [14]. The higher the cetane number, the greater the fuel's tendency to ignite. As the diesel engine dispenses with an externally supplied ignition spark, the fuel must ignite spontaneously (auto-ignition) and with minimum delay (i.e., ignition lag) when injected into the hot, compressed air in the combustion chamber. Cetane number 100 is assigned to n-hexadecane (cetane), which ignites very easily, while slow-igniting α-methyl naphthalene

is allocated cetane number 0. The cetane number of a diesel fuel is determined in a standardized CFR single-cylinder test engine with a variable compressor piston (CFR, Cooperative Fuel Research). The compression ratio of the fuel to be tested is measured at constant ignition lag. Reference fuels composed of cetane and heptamethylnonane, a component with the cetane number 15, are operated with the determined compression ratio. The proportion of cetane in the mixture is altered until the same ignition lag is obtained. The cetane proportion in percent specifies the cetane number.

The expenditure of an engine test to determine the cetane number can be avoided by determining in a constant-volume combustion chamber the ignition lag, i.e., the time interval between the start of injection and compression ignition. The derived cetane number (DCN) can be calculated from the ignition lag with the aid of empirically determined factors in accordance with the standard EN 15195 [15] or EN 16144 [16]. Alternatively, the DCN can also be determined in accordance with the standard EN 16715 [17] from the ignition lag and the combustion delay. The ignition lag is also measured in the standard EN 17155 [18]. The ignition lag of a primary reference material is used to assign the measured ignition lag to the cetane number. In this process this is referred to as the indicated cetane number (ICN).

Cetane numbers over 50 are desirable for optimum operation of modern engines, particularly under cold-starting conditions. High-quality diesel fuels contain a high proportion of paraffins with high cetane numbers. Aromatic compounds exhibit only a low ignition quality.

Cetane index

Yet another parameter of ignition quality is provided by the cetane index, which is calculated on the basis of fuel density and various points on the boiling curve. This purely mathematical parameter does not take into account the influence of cetane improvers on ignition quality. In order to limit the adjustment of the cetane number by means of cetane improvers, both the cetane number and the cetane index have been included in the list of requirements in the standard EN 590.

Boiling range

The boiling range of the fuel, i.e., the temperature range in which the fuel vaporizes and condenses again under standard conditions depends on its composition and is between 160 °C and 370 °C. Figure 4 shows boiling curves for different types of diesel fuel, the composition of which will be addressed in more detail later. A low initial boiling point is a prerequisite for a fuel suitable for use in cold weather, but also increases the risk of cavitation damage and results in greater system wear due to poorer lubrication properties.

Table 2: DIN EN 590 (October 2017): Selected requirements of diesel fuels

Requirements		Unit	Characteristic
Cetane number, min.		–	51
Cetane index, min.		–	46
Density (at 15 °C), min./max.		kg/m^3	820/845
Viscosity (at 40 °C), min./max.		mm^2/s	2.0/4.5
Sulfur content, max.		mg/kg	10
Lubricity, “wear scar diameter”, max.		µm	460
FAME content, max.		% (V/V)	7
Oxidation stability	Insoluble[3], max.	g/m^3	25
	Induction time (at 110 °C), min.	hours	20
Water content, max.		mg/kg	200
Total contamination, max.		mg/kg	24
CFPP[1] in six seasonal classes, max.[2]		°C	+5 to –20
Flash point		°C	>55

[1] Filterability limit. [2] Determined nationally, for Germany 0 to –20 °C.
[3] Sum total of insoluble substances after aging.

If, however, the end of boiling is at the high end of the temperature scale, this can result in higher soot emissions and nozzle coking. This refers to the formation of deposits as a result of the chemical decomposition of non-volatile fuel components in the spray hole and on the nozzle cone and the addition of combustion residues. The processes of carbon depositing are highly complex. Above all non-volatile fuel components from the refinery cracking processes contribute to coking.

The degree of coking, which is calculated from the measured hydraulic throughflow of the uncleaned and the cleaned nozzle, is for a specified nozzle geometry a measure of the tendency of a fuel to form deposits in the injection nozzles. Conversely, it is possible to assess with a reference fuel how the design of different nozzle configurations influences the formation of carbon deposits.

When the end of boiling is higher, the ingress of fuel over the cylinder walls into the engine oil is greater. For this reason, the percentage of non-volatile fuel components should not be too high.

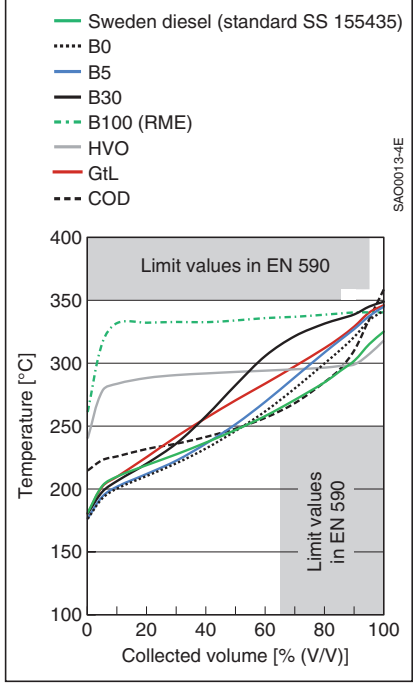
Figure 4 shows the boiling curves of different types of diesel fuels:

– Sweden diesel, fuel with reduced aromatic-compound content in accordance with the standard SS 155435 [19],
– B0, petroleum-based diesel fuel,
– B5, petroleum-based diesel fuel with 5 % biodiesel,
– B30, petroleum-based diesel fuel with 30 % biodiesel,
– B100, pure biodiesel, RME (Rapeseed Methyl Ester),
– HVO, paraffinic diesel fuel through hydrogenation of vegetable oils (Hydro-treated Vegetable Oil),
– GtL, paraffinic diesel fuel, synthetically produced from natural gas (Gas-to-Liquid),
– COD, diesel fuel from olefinic refinery components (Conversion-of-Olefins-to-Distillates).

The collected volume of condensate is plotted against the temperature in Figure 4. The high boiling range (320 to 360 °C) of biodiesel (B100) compared with petroleum-based conventional diesel fuel (B0) can be discerned. The admixture of 5 or 30 % biodiesel (B5 or B30) already impacts significantly on the higher content of the boiling range. The limitation of the admixture of biodiesel to max. 7 % (V/V) in EN 590 is not least due also to its non-volatility. Purely paraffinic fuels are more favorable than biodiesel with regard to the boiling range. Diesel fuel synthetically produced from natural gas (Gas-to-Liquid, GtL) exhibits similar boiling characteristics to conventional diesel fuel or Swedish diesel fuel reduced in aromatic-compound content in accordance with the standard

Figure 4: Boiling curves of different types of diesel fuels

B0, B5…B100 Admixtures of biodiesel,
RME Rapeseed methyl ester,
HVO Hydro-treated vegetable oil,
GtL Gas-to-Liquid,
COD Conversion-of-Olefins-to-Distillates.



SS 155435. In the boiling curve of paraffinic fuel created by the hydrogenation of vegetable oils (Hydro-treated Vegetable Oil, HVO) the relatively uniform composition in the middle boiling range becomes clear because the paraffins contained in the HVO have come from a narrowly limited fatty-acid spectrum.

Filterability limit
Precipitation of paraffin crystals at low temperatures can result in fuel-filter blockage, ultimately leading to an interruption of the fuel flow. In worst-case scenarios, paraffin particulates can start to form already at temperatures of 0 °C or even higher. The cold resistance of a fuel is assessed by means of the temperature at which the fuel filter becomes clogged under defined test conditions (CFPP, Cold Filter Plugging Point, filterability temperature limit). Standard EN 590 defines different limit values for the CFPP in various classes, and can be defined by individual member states depending on the prevailing geographical and climatic conditions.

Formerly, owners sometimes added regular gasoline to their vehicle fuel tanks to improve the cold response of diesel fuel. This practice is no longer necessary now that fuels conform to standards, and this may in any case cause damage particularly in today's widely used systems with high-pressure fuel injection.

Flash point
The flash point is defined as the temperature at which the quantities of vapor which a combustible fluid emits to the atmosphere are sufficient to allow a spark to ignite the air/vapor mixture above the fluid. For safety reasons (e.g., for transportation and storage), diesel fuel must comply with Hazard Class A III, i.e., its flash point must be above 55 °C. Less than 3 % gasoline in the diesel fuel is sufficient to lower the flash point to such an extent that ignition becomes possible at room temperature.

Density
The energy content per unit of volume increases in the case of purely fossil diesel fuel produced only from crude oil with density. Assuming constant activation of the injectors (i.e., constant injected fuel quantity), the use of fuels with widely different density causes variations in mixture ratios (change in the air/fuel ratio λ) due to fluctuations in calorific value. When an engine runs on fuel that has a high type-dependent density, engine performance and soot emissions increase; at low fuel density, these parameters drop. For this reason, the type-dependent density spread for diesel fuel is tightly limited.

When biodiesel is admixed with diesel fuel, the density of the mixture increases due to the higher density of biodiesel, but the energy content decreases due to the amount of oxygen contained.

Viscosity
Viscosity is a measure of a fuel's resistance to flow due to internal friction. Excessively low viscosity of the diesel fuel leads to increased return quantities in the injector, as a result of which the fuel-injection system is additionally heated up and in some cases the injected fuel quantity required for full load is no longer achieved. In addition, the risk of wear and cavitation erosion increases.

Much higher viscosity – for instance when pure biodiesel (FAME) is used – causes a higher peak injection pressure at high temperatures in non-pressure-regulated systems (e.g., unit injector systems) compared with petroleum diesel. The fuel-injection system may not therefore be applied with petroleum diesel to the permissible peak pressure. High viscosity also changes the spray pattern due to the formation of larger droplets. The droplet distribution in the combustion chamber is called the spray pattern. Since the fuel viscosity increases sharply in cold conditions, restrictions in fuel delivery may be experienced as a result of the reduction of the volumetric flow.

Lubricity
The hydrodynamic lubricity of diesel fuels is not as important as the lubricity in the mixed-friction range. The introduction of environment-compatible, hydrogenation-desulfurized fuels resulted in huge wear problems with distributor injection pumps in the field. Desulfurization also sees the removal of fuel components which are important to lubricity. Lubricity enhancers have to be added to many fuels to avoid these problems. The standard EN 590 prescribes a minimum lubricity, measured as wear of max. 460 µm in an oscillation wear test (High-Frequency Reciprocating Rig, HFRR [20]).

The minimum-lubricity requirement is also achieved when sufficient biodiesel, usually at least 4 % (V/V), is admixed with the diesel fuel.

Oxidation stability
The constituents of diesel fuels are subject to a more pronounced extent than gasoline to oxidation caused by oxygen in the air. This is the case particularly when biodiesel is admixed. Polymers and acids are produced during oxidation, which is also referred to as fuel aging. Aging polymers can directly give rise to the formation of coatings and to gumming in the fuel-injection system. Aging acids frequently react with metal ions to produce organic salts and soaps (carboxylates), which are insoluble in diesel fuel and therefore also form coatings.

It is important that the fuel, while it is in the vehicle, has a sufficient aging reserve. The aging reserve is measured as induction time under the test conditions defined in EN 15751 [21] at 110 °C. When diesel fuel at the filling-station pump exhibits an aging reserve of at least 20 hours according to this test standard, it is ensured that the fuel is stable enough for normal use. In the past, failures caused by fuel aging have only occurred in isolated cases when vehicles were subject to very long idle periods, in special agricultural applications, or also in the construction sector.

Total contamination
Total contamination refers to the sum total of undissolved foreign particulates in the fuel, such as sand, rust, and insoluble organic components, including aging polymers of the fuel. The standard EN 590 permits a maximum total contamination of 24 mg/kg. Very hard silicates which occur in mineral dust are specially damaging to high-pressure fuel-injection systems with narrow gap widths. Even a fraction of the permissible overall contamination level if the particulates were abrasive would cause erosive and abrasive wear.

Since 2014 there have in many regions of Europe been a large number of reports of damage to fuel-injection systems caused by abrasive particulates (e.g., to the seat of solenoid valves) [22]. Wear of this nature results in valve leakage, which lowers fuel-injection pressure and engine performance, and increases particulate emissions. Typical European diesel fuels contain about 2,500 particulates per 1 m*l*. Particulate sizes around 4 µm are particularly critical. High-performance fuel filters with high filtration efficiency are therefore necessary to prevent damage caused by particulates.

Water in diesel fuel
Up to approximately 100 mg/kg water dissolve in diesel fuel at room temperature. The solubility limit depends on the composition of the diesel fuel, particularly the aromatic-compound content, the additives, the proportion of biodiesel, and the ambient temperature. The standard EN 590 permits a maximum water content of 200 mg/kg. Although diesel fuels in storage tanks can contain much higher amounts of water, market surveys of fuels show that water content rarely exceeds 200 mg/kg. Samples often do not detect any water, or detection is incomplete, since water is deposited on walls in a separate phase in the form of undissolved, “free” water, or it settles at the bottom. Whereas dissolved water does not damage the fuel-injection system, even very small quantities of free water can cause wear or corrosion damage to fuel-injection components within a short period of time.

Additives
Additives, long a standard feature in gasolines, have attained increasing significance as quality improvers in diesel fuels. Additive packages, which are effective in a number of respects, are used for the most part. As the total concentration of the additives generally lies below 0.1 %, the fuel's physical characteristics – such as density, viscosity, and boiling curve – remain unchanged.

Lubricity enhancers
It is possible to improve the lubricity of diesel fuels with poor lubrication properties, caused, for example, by hydrogenation processes during desulfurization, by adding fatty acids or glycerides. Glycerides are partially cracked vegetable oils. Even biodiesel contains proportions of glycerides. In this case, if diesel fuel already contains biodiesel, no further lubricity enhancers are added.

Cetane improvers
Nitric-acid esters of alcohols, which shorten the duration until ignition (ignition lag), are frequently used as cetane improvers. They help, particularly during cold starting, to prevent an increase in combustion noise (engine noise) and extreme smoking.

Flow improvers
Flow improvers consist of polymer substances that lower the filterability temperature limit. They are generally added in winter to ensure trouble-free operation at low temperatures. Although flow improvers cannot prevent the precipitation of paraffin crystals from diesel fuel, it can severely limit their growth. The crystals produced are then so small that they do not clog the filter pores completely and allow the passage of fuel until they melt again as the engine heat begins to build up.

Detergents
Detergents are cleaning additives which are added to the fuel to keep nozzle holes free from deposits on the inside and at the outlet. The cleaning effect of detergents is based on their molecular structure of groups of different polarity: Long nonpolar hydrocarbon chains ensure good solubility in the fuel while coordination to the for the most part also polar coatings is facilitated by way of polar groups. Detergents function in this way as carriers with which existing deposits can be removed via the fuel. Furthermore, these additives can also prevent the build-up of new deposits by coordination of the polar groups to metal surfaces.

Corrosion inhibitors
Typical corrosion inhibitors are longer-chain organic monomeric and dimeric acids, amines or ammonium salts. These compounds are deposited with their polar end on the surfaces of metal parts and protect them against corrosion if water is entrained.

Antifoaming agents (defoamers)
Adding defoamers helps to avoid excessive foaming when the vehicle is refueled quickly. Antifoaming agents such as silicones reduce the surface tension, causing bubbles that are created to burst rapidly.

Blend components for fossil diesel fuels

Biodiesel

At present, biodiesel is the most important alternative fuel for diesel engines. The term biodiesel covers fatty acid esters which are created through transesterification of oils or greases with methanol. This creates fatty acid methyl esters (FAME). Fatty acid methyl esters are in their molecule size and thus also in their physical-chemical properties much more similar to diesel fuel than vegetable oils. Biodiesel and vegetable oils are therefore not the same. Nevertheless, biodiesel differs significantly from petroleum diesel, since the fatty acid esters are polar and chemically reactive. Conventional diesel fuel on the other hand is a nonpolar mixture of paraffins and aromatic compounds and therefore less active.

Manufacture

Vegetable oils or animal fats can be used as the starting material for biodiesel. In Europe primarily rape oil is used, in North and South America soybean oil, in Asia palm oil, and on the Indian subcontinent to an even smaller extent the oil from the purgier nut (jatropha). Used frying oil methyl esters (UFOME) are produced worldwide. Because of the global trade in biodiesel and its raw materials, fuels containing FAME as a rule contain mixtures from different sources.

Because transesterification of grease and oils is technically easier with methanol than with ethanol, the methyl esters are manufactured almost exclusively. Methanol is generally produced from coal. Therefore, fatty acid methyl esters cannot, strictly speaking, be seen as fully biogenic.

Properties

The properties of biodiesel are determined by various factors. The various vegetable oils differ in the composition of the fatty-acid blocks and demonstrate typical fatty-acid patterns. The type and quantity of unsaturated fatty acids have, for example, a decisive influence on the cold resistance and resistance to aging of biodiesel. The properties are also affected by the pretreatment of the vegetable oils and the production process of the biodiesel.

Standards

The quality of biodiesel is regulated in fuel standards. If technically tenable, limitations with regard to raw materials must be avoided. The quality requirements for biodiesel are therefore described not by way of the raw-material composition, but predominantly by the material properties. It is essential in particular to ensure suffi-

Table 3: DIN EN 14214 (May 2019): Selected requirements of FAME

Requirements	Unit	Characteristic
Density (at 15 °C), min./max.	kg/m^3	860/900
Viscosity (at 40 °C), min./max.	mm^2/s	3.5/5.0
Sulfur content, max.	mg/kg	10
Content of alkali metals (Na + K), max.	mg/kg	5.0
Content of alkaline earth metals (Ca + Mg), max.	mg/kg	5.0
Content of total glycerin, max.	% (m/m)	0.25
Acid number, max.	mg KOH/g	0.50
Oxidation stability (induction time at 110 °C), min.	hours	8
Water, max.	mg/kg	500
Total contamination, max.	mg/kg	24
CFPP[1], in six seasonal classes in each case, max.		
Pure fuel	°C	+5…–26[2]
Blend component	°C	+13…–10[3]
Flash point	°C	>101

[1] Filterability limit.
[2] Determined nationally, for Germany 0 to –20 °C.
[3] Determined nationally, for Germany 0 to -10 °C.

cient cold resistance, good resistance to aging (oxidation stability), and to eliminate contamination caused by the process.

The European standard EN 14214 is the most internationally comprehensive specification for biodiesel and describes the requirements relating to the use of biodiesel as a pure fuel and as a blend component for diesel fuel. This standard serves to ensure that the quality of biodiesel meets the requirements in both applications (Table 3).

The American biodiesel standard ASTM D6751 is less quality-oriented. For example, the minimum requirement for oxidation stability in this standard is less than half the aging reserve (induction time) specified in EN 14214. Thus, in the USA the risk of problems arising as a result of fuel aging, in particular under limit-value application and field conditions, is higher.

Other countries such as Brazil, India and Korea have geared themselves, to a large extent, towards the European B100 standard EN 14214.

Use in vehicles

Since the tax benefit was discontinued biodiesel has barely been used as a pure fuel (B100) in Germany. The introduction of diesel particulate filters and stringent requirements regarding the limitation of vehicle emissions under field conditions even over high mileages limit the use of B100 to a few special applications. Originally B100 was used predominantly in commercial vehicles, because the annual high mileage ensured fast consumption, which enabled problems with insufficient oxidation stability to be avoided.

From an engine-engineering viewpoint, it is more favorable to use biodiesel as a low-percentage mixture with petroleum diesel. The high proportion of petroleum diesel ensures sufficient stability and, at the same time, the biodiesel ensures an improved lubricating effect.

For practical purposes, it is important to specify not only the pure component B100, but also the diesel/biodiesel mixtures offered on the market. Here, the trend is towards admixtures up to max. 7 % biodiesel (B7 in Europe).

In closed fleets, higher proportions of biodiesel are also used (B30 in France, B20 in the USA). In the case of higher contents, the high boiling point of biodiesel can cause it to be heavily introduced into the engine oil after its has been injected into the combustion chamber via condensation on the cylinder walls. This affects, above all, vehicles which are fitted with diesel particulate filters, and in which filter regeneration occurs by way of a retarded secondary injection. Depending on the application, it is possible, particularly in slow part-load operation, to encounter an unacceptably high introduction of biodiesel, which necessitates shorter oil-replacement intervals. The higher boiling point of biodiesel can be taken into account by application in new developments of engines and vehicles.

An additional standard has been developed in Europe for B20/B30, EN 16709 [23], in which two grades with different biodiesel content are specified: B20 with 14 to 20 % and B30 with 24 to 30 % biodiesel. To avoid problems caused by fuel aging, the minimum oxidation-stability requirement from EN 590 (B7) has been adopted for both grades, i.e., the induction time must be at least 20 hours, irrespective of the biodiesel content. In this way, the new European standard EN 16709 differs fundamentally from the American standard ASTM D 7467 [24], in which merely an induction time of at least 6 hours was laid down for biodiesel contents of 6 to 20 %.

The standard EN 16734 [25] has also been adopted in Europe; this standard covers diesel fuels with biodiesel contents of 0 to 10 % and thus overlaps to a large extent with the standard EN 590 for standard diesel fuel (0 to 7 % biodiesel). Still the market demand for this additional quality is low. Furthermore, this fuel is not permitted to be placed on the market in all European countries.

Vegetable oils

Vegetable oils which have not been transesterified into biodiesel may not be admixed with diesel fuels due to their high density, high viscosity, and low volatility. Previously, vegetable oils, particularly rapeseed oil, were used with great success in older diesel engines subject to minimal emission requirements. Only few

modern diesel engines with high-pressure fuel injection can be run on vegetable oils, for example in tractors which have been approved for such operation.

Paraffinic diesel fuels
Further blend components for fossil diesel fuels are purely paraffinic fuels, which consist exclusively of saturated hydrocarbons. Since there are no aromatic components, the particulate, HC and CO emissions of purely paraffinic fuels are significantly reduced when compared with conventional fuels.

Manufacture
Paraffinic fuels can be created in three different ways:
– Fischer-Tropsch process [26]
– Hydrogenation of vegetable oils
– COD process (Conversion of Olefins to Distillates)

Fischer-Tropsch process
The starting product required is synthesis gas, which consists of hydrogen (H_2) and carbon monoxide (CO). It can be created from natural gas, coal or biomass. By converting synthesis gas, it is possible to build up on catalysts linear, straight-chain hydrocarbons (n-paraffins). The Fischer-Tropsch catalysts function quite non-specifically such that a wealth of different components is obtained, starting with gases through short-chain gasoline components, kerosene and diesel paraffins, right down to oils and waxes of high molecular weight. For reasons of economy, the production mixture is for the most part optimized to a maximum diesel yield. The frequently used designation synthetic diesel fuels makes clear that the hydrocarbons in the Fischer-Tropsch process are made from the building blocks H_2 and CO.

These fuels were originally also referred to as designer fuels, because the notion existed that the composition of synthetic diesel fuel could be geared exactly to the demands of diesel-engine technology. In view of the wide range of products obtained from Fischer-Tropsch synthesis, the notion of producing fuels of customized composition does not appear realistic.

For the designation of this type of synthetic fuels it is common practice to use the terms derived from the conversion process Gas-to-Liquid (GtL), Coal-to-Liquid (CtL), or Biomass-to-Liquid (BtL), depending on whether the paraffins have been created from natural gas, coal or biomass.

The production of CtL and GtL is economically significant. The production of GtL is only worthwhile with large natural-gas deposits where the natural gas cannot be attributed to any direct use. CtL and GtL are based on fossil energy sources such that no reduction in CO_2 emissions is achieved. In the case of BtL, there is no CO_2 advantage. However, industrial plants that produce BtL are not as yet in operation.

Biomass is utilized completely in the manufacture of BtL. The plants are first broken down into chemical building blocks from which fuels are then synthesized in a subsequent step. Fuels which are manufactured in this way are called second-generation fuels. This procedure differs fundamentally from the hitherto customary methods, which are used for the production of first-generation fuels and are based on the use of agricultural crops. Biodiesel can be created by transesterification and bioethanol by fermentation from the components contained therein such as fats, starches and sugars.

Hydrogenation of vegetable oils
Paraffinic diesel fuels can also be obtained by hydrogenating fats and oils. The term hydro-treated vegetable oil (HVO) is often used for these products. Unlike transesterification into biodiesel, conversion with hydrogen places fewer demands on the origin and quality of the starting materials. Hydrogenation results in a cracking of fats and oils, during which all the oxygen atoms and unsaturated bonds are also removed. Long-chain paraffins are obtained from the fatty acids, the glycerin content is converted into propane gas, and the oxygen is bound as water. Because paraffins are created from biomass in this way, they are referred to as bioparaffins. The production of bioparaffins can be implemented in separate plants or even integrated into existing refinery processes.

The hydrogenation of vegetable oils to HVO is developing increasingly into a significant alternative to the transesterification of biodiesel.

COD process
The third way of producing paraffins is to convert olefins in accordance with the COD process (Conversion of Olefins to Distillates), frequently only as a subsequent step to previous refinery processes. Olefinic product fractions are used for this purpose. Paraffins can be produced by oligomerization, a specifically controlled reaction between individual olefins, and hydrogenation.

Reprocessing
All three manufacturing processes described here are followed by an additional isomerization step in which linear molecules are converted into branched molecules. This is the only way to achieve the required cold resistance.

Properties
Regardless of the type of production, paraffinic hydrocarbon mixtures with very similar chemical compositions and excellent engine-engineering properties are created. The fuels are sulfur- and aromatic-free and have partly high cetane numbers. But because their density is below the lower limit value defined in EN 590, the new specification EN 15940 [27] was developed (Table 4). This standard describes the quality which is required to use the fuel as a pure component.

Use in vehicles
Paraffinic fuel in accordance with EN 15940 can be used in vehicles which have been approved for this fuel. Pure paraffinic fuels are used mostly in older vehicles with no diesel particulate filters and lend themselves to use above all in centers of population in order to reduce the particulate load on a local level. To reduce the CO_2 emissions of vehicle fleets, HVO has recently been increasingly used in the haulage industry.

Paraffinic hydrocarbons are ideal for marketing as blend components in premium diesel fuels. Furthermore, diesel fuels which fail to satisfy the requirements established in EN 590 can be improved by the addition of paraffinic components to such an extent that they conform to the standard.

Table 4: DIN EN 15940 (October 2019): Selected requirements of paraffinic diesel fuels

Requirements		Unit	Characteristic
Cetane number, min.	Class A	–	70
	Class B	–	51
Density (at 15 °C), min./max.	Class A	kg/m^3	765/800
	Class B	kg/m^3	780/810
Viscosity (at 40 °C), min./max.		mm^2/s	2.0/4.5
Sulfur content, max.		mg/kg	5
Total aromatic content (including polyaromatics), max.		% (m/m)	1.1
Lubricity, “wear scar diameter”, max.		µm	460
FAME content, max.		% (V/V)	7
Water, max.		mg/kg	200
Total contamination, max.		mg/kg	24
CFPP [1] in six seasonal classes, max.[2]		°C	+5 to –20
Flash point		°C	>55

[1] Filterability limit.
[2] Determined nationally, for Germany 0 to –20 °C.

Methane

Methane (CH_4) is the simplest aliphatic hydrocarbon. Gasoline and diesel engines can be run on methane.

Natural gas

Composition

Natural gas is fossil in nature. Natural gas is promoted worldwide and can be prepared with relatively little effort. The main constituent of natural gas is methane (CH_4) with a minimum content of 80 % (V/V). Further constituents are inert gases, such as carbon dioxide, nitrogen and short-chain hydrocarbons. Oxygen and hydrogen are also contained. Depending on its origin, however, its composition varies, which results in fluctuations when it comes to density, calorific value and knock resistance. The way in which natural gas is transported also impacts on its composition.

Unlike when it is transported as gas through pipelines, natural gas that is distributed by ships in liquid form as LNG (liquefied natural gas) contains hardly any CO_2, pentane, hexane, and hydrogen sulfide. The properties of natural gas as a fuel are set out in the European standard EN 16723-2 [30], in which standardized quality requirements with regard to natural gas, to pure biomethane and to blends of natural gas and biomethane are defined.

Table 5 provides an overview of the composition and properties of natural gas worldwide.

To be able to better describe the combustion behavior of gaseous fuels like natural gas, two additional product properties are defined.

Methane number

The methane number describes the knock properties and specifies the volume

Table 5: Composition and properties of different natural-gas qualities ([28], [29])

Parameter	Symbol	Unit	Russia Group H	North Sea Group H	Denmark Group H	Libya LNG	Nigeria LNG	Egypt LNG	Biomethane	Biomethane + LPG	Mix Group H	Holland Group L	Natural-gas network Group L	Weser/Ems Group L
Methane	CH_4	mol%	96.96	88.71	90.07	81.57	91.28	97.70	96.15	90.94	87.39	82.70	85.41	87.78
Nitrogen	N_2	mol%	0.86	0.82	0.28	0.69	0.08	0.08	0.75	0.69	1.60	10.32	9.24	9.26
Carbon dioxide	CO_2	mol%	0.18	1.94	0.60				2.90	2.68	1.94	1.57	1.63	2.33
Ethane	C_2H_6	mol%	1.37	6.93	5.68	13.38	4.62	1.80			8.00	4.33	2.99	0.57
Propane	C_3H_8	mol%	0.45	1.25	2.19	3.67	2.62	0.22		5.00	0.90	0.78	0.51	0.04
n-butane	nC_4H_{10}	mol%	0.15	0.28	0.90	0.69	1.40	0.20		0.50	0.14	0.22	0.15	0.02
n-pentane	nC_5H_{12}	mol%	0.02	0.05	0.22						0.02	0.06	0.03	
n-hexane	nC_6H_{14}	mol%	0.01	0.02	0.06						0.01	0.05	0.04	
Hydrogen	H_2	mol%												
Oxygen	O_2	mol%							0.20	0.19				
Sum		mol%	100	100	100	100	100	100	100	100	100	100	100	100
Calorific value Hs on a volumetric basis	H_S	MJ/m³	40.3	41.9	43.7	46.4	44.0	40.7	38.3	41.9	41.5	37.2	37.2	35.4
Calorific value Hs on a volumetric basis	H_S	kWh/m³	11.2	11.6	12.1	12.9	12.2	11.3	10.6	11.6	11.5	10.3	10.3	9.8
Rel. density (air ≙ 1)	d	–	0.574	0.629	0.630	0.669	0.624	0.569	0.587	0.641	0.631	0.647	0.632	0.619
Wobbe index	W_S	MJ/m³	53.1	52.9	55.0	56.7	55.7	53.9	50.0	52.3	52.3	46.3	46.3	44.8
Wobbe index	W_S	kWh/m³	14.8	14.7	15.3	15.8	15.5	15.0	13.9	14.5	14.5	12.9	12.9	12.5
Methane number	MN	–	92	79	73	65	71	92	103	77	80	86	90	102

percentage of methane in a methane/hydrogen mixture which exhibits in a test engine under standard conditions the same knock characteristics as the gaseous fuel to be tested. Pure methane is assigned the methane number 100, hydrogen the methane number 0. As with the definition of the octane number for gasolines, the following applies to gases and gas mixtures: the lower the methane number, the lower the knock resistance.

Wobbe index

The upper Wobbe index W_s specifies the gross calorific value H_s on a volumetric basis. For many applications it is important to know the energy content of the volumetric flow. Only with reference to the Wobbe index can the flow of gas be set to the desired amount of energy and the gas quality easily corrected in the event of changes. A consideration of the calorific value alone is not enough. The upper Wobbe index W_s is obtained by dividing the calorific value H_s by the square root of the relative density of the combustion gas. The relative density of the combustion gas is the quotient of the density of the combustion gas and the density of dry air, both determined under the same reference conditions (pressure and temperature). If, however, all that matters is the mechanically utilizable energy content in the volumetric flow, reference is made to the lower Wobbe index W_i, i.e., the net calorific value H_i on a volumetric basis.

Biomethane

Biomethane is treated biogas which is made from biogenic materials such as energy crops, liquid manure or waste containing biomass. The CO_2 savings from the use of biogenic basic material can be set off against the CO_2 emissions during combustion in such a way that the total CO_2 emissions for biomethane are clearly reduced compared with fossil natural gas. Production volumes of biomethane which exceed the local demand are frequently fed into the natural-gas network. In summer, when the demand for natural gas is too low for heating purposes, natural gas can contain significant amounts of biomethane.

A lot of CO_2 is produced in biogas plants as well as methane. Processes at plant scale in which CO_2 can be biologically converted by highly specialized microorganisms into methane are well advanced and are already being introduced occasionally.

Synthetic methane

In initial plants methane is also obtained from electricity whereby hydrogen (H_2) is generated by electrolysis from water. The conversion of H_2 with CO_2 to CH_4 is then performed chemically in the presence of catalysts in accordance with the Sabatier process.

Storage of natural gas

Natural gas is stored either in gas form (CNG, compressed natural gas) at a pressure of 200 bar or as a liquefied gas (LNG, liquefied natural gas) at –162 °C in a cold-resistant tank. LNG requires only one third of the storage volume of CNG, however the storage of LNG requires a high expenditure of energy for liquefying. Filling stations dispense natural gas almost exclusively in the form of CNG. LNG is as a rule only offered if the filling station can already be supplied with LNG.

CO_2 emissions

The hydrogen/carbon ratio of natural gas is roughly 4:1, that of gasoline on the other hand is 2,3:1. As a result of the lower amount of carbon in natural gas, it produces less CO_2 and more H_2O than gasoline when it is burned. A gasoline engine converted to natural gas creates, for a comparable power output without any further optimization, approximately 25 % fewer CO_2 emissions than when run on gasoline. However, if a small methane slip remains during combustion, the additional methane emissions must be taken into account in the climate-damaging-potential footprint since direct methane emissions have a 20-times greater impact than CO_2. Even escaping gaseous methane is relevant to the carbon footprint when for example an overpressure builds up in the LNG storage tanks and this is discharged to the atmosphere.

Use in vehicles

Natural gas is currently used primarily in passenger cars with gasoline engines which are of bivalent design, i.e., can be run separately on natural gas and gasoline. For buses and trucks there are gasoline gas engines which are of monovalent design for natural gas [31]. Also used in the diesel gas engine, in which natural gas in burned in mixed operation with diesel fuel. The natural gas is in this cased injected separately from the diesel fuel with the air into the engine [32].

Alternatively, natural gas can also be injected directly in a pilot-injection process together with the diesel fuels [33]. Whether the market for natural-gas vehicles will gain in importance depends primarily on the purchase and running costs for the vehicle owner, but also on the expansion of the filling-station infrastructure to CNG and LNG.

Liquefied petroleum gas

Liquefied petroleum gas (LPG), also called liquid gas, is obtained during the extraction of crude oil and during different refinery processes. LPG is a mixture of the main components of propane and butane. It can be liquefied at room temperature under comparatively low pressure. Because it has a lower carbon content than gasoline, approximately 10 % less CO_2 is produced during combustion. The octane number is approximately 100 – 110 RON. The demands placed on LPG for use in motor vehicles are laid down in the European standard EN 589 [34].

Passenger cars for liquefied petroleum gas are also for the most part bivalent and can therefore be run separately on gasoline and liquefied petroleum gas. However, liquefied petroleum gas can cause significant wear on the intake valves if valve or seat-insert materials have not been adapted to running on LPG.

Hydrogen

Manufacture

Hydrogen (H_2) can be produced from fossil or renewable sources. The classic production methods from fossil raw materials such as steam reforming of natural gas, gasification of coal in the presence of water or partial oxidation of mineral oil result in the release of considerable amounts of CO_2. Although no CO_2 emissions are produced locally during the engine combustion of hydrogen, the use of hydrogen of a fossil nature does not inevitably lead to a CO_2 advantage in direct comparison with natural gas, gasoline or diesel fuel.

Hydrogen can be regeneratively manufactured both by pyrolysis of biomass and by electrolysis or solar-thermal processes directly from water. CO_2 savings are obtained as a result of the production process and the overall expenditure of energy and are greater the more regeneratively produced energy is used in the process.

Because hydrogen today is primarily obtained industrially from natural gas, processes are also being developed with the aim of capturing the CO_2 created and injecting it into repositories (CCS, carbon capture and storage).

Storage

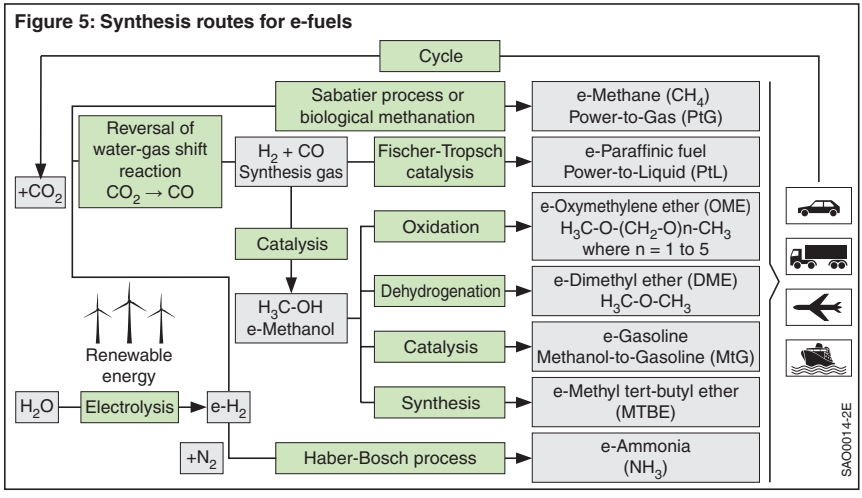
Hydrogen may have a very high mass-related energy density (approximately 120 MJ/kg and thus almost three times as high as that of gasoline), but its volume-related energy density is very low on account of the low specific density. When it comes to storage, this means that the hydrogen has to be compressed either under pressure (at 350 – 700 bar) or by liquefaction (cryogenic storage at –253 °C) in order to achieve an acceptable tank volume.

Another possibility is for the hydrogen to be stored as a hybrid. Certain metals and alloys form, with hydrogen, metal hydrides. The hydrogen that is embedded in the metal's atomic lattice can be released again by heating.

Use in vehicles

Hydrogen can be used both in fuel-cell drives and directly in internal-combustion engines. In the long term, emphasis is expected to be placed on use in fuel cells because this achieves better efficiency than in hydrogen internal-combustion engines.

Figure 5: Synthesis routes for e-fuels



Electricity-generated regenerative fuels

E-fuels refer to synthetic fuels which are manufactured with the aid of regeneratively produced electricity.

Manufacture
In the first step electrolysis is used to break down water into hydrogen (H_2) and oxygen (O_2). The created hydrogen, e-H_2, is the simplest e-fuel, which can be used directly, for example in a fuel cell.

Figure 5 shows how carbon-containing e-fuels can be constituted from e-H_2 in secondary reactions. Methane, e-cH_4, can be created by converting e-H_2 with CO_2. Gaseous e-fuels are called Power-to-Gas (PtG). It goes without saying that the CO_2 used must also be of a regenerative nature. Regenerative CO_2 can derive for example from biomass or be obtained from the ambient air. Due to the higher concentration of CO_2 industrial waste gases, for example from steel or cement works, suggest themselves as sources, but this CO_2 is not regeneratively produced.

To obtain in particular liquid e-fuels, accordingly named Power-to-Liquid (PtL), synthesis gas, a mixture of e-H_2 and carbon monoxide, is required. Carbon monoxide is available from CO_2 by a reversal of the water-gas shift reaction.
Starting out from synthesis gas, two basic reaction paths are possible. On the one hand, linear chains of hydrocarbons can be built up by means of Fischer-Tropsch synthesis. They are, depending on the chain length, typical constituents of gasoline, kerosene or diesel fuel.

Alternatively, e-methanol can also be produced from synthesis gas. Because limits are imposed on the direct use of methanol in gasoline engines, its further conversion to MTBE (methyl tertiary-butyl ether) or MtG gasoline (Methanol-to-Gasoline) suggests itself. Due to its high volatility and low cetane number of 5, methanol is not a fuel component for diesel engines Converting to dimethyl ether (DME) or to the higher oxymethylene ethers (OME) produces fuels with a cetane number > 70.

Also being discussed as a carbonless e-fuel aside from e-H_2 is ammonia (NH_3), which is easily available in the Haber-Bosch process from e-H_2 and nitrogen (N_2) and is possible both as a H_2 store and as a fuel.

Use in vehicles
All e-fuels which do not differ from the corresponding fossil or biogenous components or are very similar qualitatively to these are called drop-in fuels. e-H_2, e-CH_4, e-MTBE, MtG and Fischer-Tropsch gasoline and Fischer-Tropsch diesel fuel can either be used directly at 100 % in the overall vehicle population or else at least replace the conventional components in large if not equal proportions. This substitution potential, which would entail additional saving of significant quantities of CO_2 in the traffic sector, can from a technical standpoint be used directly and depends on the availability of e-fuels.

Other e-fuels such as e-DME or e-OME are, on account of their properties, not compatible with the vehicles in the on-the-road vehicle population. A new development of engines or at least their comprehensive adaptation is required to accommodate these fuels. The particulate emissions of these fuels are greatly reduced. In part, there can be technical advantages in the design of exhaust-gas treatment. Because of the necessary changes, however, these types of e-fuel are expected to find use in special applications and less so in the larger-scale vehicle population.

Prospects for e-fuels
Although the production of e-fuels requires an additional energy expenditure which is above its own energy content and the costs of their production, depending on the number of synthesis steps, far exceed the costs of the fossil components, e-fuels will be afforded opportunities in the future. Without e-fuels, for which a significant reduction of the production costs is forecast, the long-term CO_2 targets in the traffic sector cannot be achieved. Militating for the use of e-fuels is furthermore the fact that purely battery-electric drives can only gradually replace the internal-combustion engine in heavy goods vehicle traffic and long-distance transportation in the foreseeable future. For ships and in particular for aviation it is up to now

difficult to imagine that there could be a technical alternative to the replacement of fossil fuels with biofuels and e-fuels.

Research projects have hitherto failed to bring onto the market considerable quantities of e-fuels for the transport sector. Because the construction of large-scale plants for the production of e-fuels requires long lead times for planning, financing, design, and upscaling it is hardly expected by 2030 that significant proportions of fossil fuels can be replaced with e-fuels. All the more important then are short-term political moves that create a solid framework for investors so that e-fuels can make a contribution to meeting the CO_2 targets in traffic applications.

Results of research work into e-fuels and scenarios regarding their introduction are the subject of numerous publications. In the meantime, handbooks provide a good overview of this subject area [35].

Ethers

Dimethyl ether

Dimethyl ether (DME) with the structure $H_3C\text{-}O\text{-}CH_3$ has a boiling point of –25 °C and must therefore be kept under pressure at room temperature so that it can be injected in liquid form. The two methyl groups are separated by an oxygen atom in the DME. Molecules with carbon atoms isolated from each other generate little soot during combustion. DME also has a high cetane number, reducing combustion and lowering NO_x emissions. The high oxygen content of DME is however associated with a low calorific value H_u of 28.8 MJ/kg and thus a high volumetric increased consumption compared with diesel fuel (H_u of 43.0 MJ/kg).

Although DME from a combustion perspective is an ideal fuel for diesel engines, in reality only a few specially adapted vehicles use this fuel. DME liquefied under pressure is, due to its high vapor pressure and its low viscosity, not compatible with the existing injection components and engine designs developed for conventional diesel fuels.

The use of DME will, due to the necessary additional technical outlay on the vehicle, remain limited to just a few fleet vehicles in niche markets. Nevertheless, the possibility of using DME has been the subject of discussion again of late in connection with e-fuels. The manufacture of DME is performed by means of synthesis gas (H_2 and CO) is currently based mainly on carbon or natural gas, supplemented to a small extent by biomass or waste materials. If in future synthesis gas can be produced in larger quantities from hydrogen from electrolysis e-H_2 and regenerative CO_2, this could drive forward the use of DME in engine applications.

Oxymethylene ethers

Oxymethylene ethers (OME) too have carbon atoms isolated from each other by oxygen bridges and are therefore of interest due to their lower tendency to sooting.

It is clear from their chemical structure $H_3C\text{-}O\text{-}[CH_2\text{-}O]_n\text{-}CH_3$ that with OME ether molecules of different chain length are involved. The base component with $n=1$ (OME1) and a mixture of the longer

ethers with n = 3 to 5 (OME3-5) are being discussed as possible future fuels. OME too is an e-fuel when the synthesis gas from which OME is produced under other reaction conditions like DME consists of hydrogen from electrolysis and regenerative CO_2.

OME1

OME1 is technically easily available from synthesis gas. This oxymethylene ether, which is chemically more accurately called dimethoxymethane, is however not suitable either as a pure substance or as a blend component for existing diesel engines, although OME1 unlike DME is still liquid at room temperature.

With a low boiling point of 42 °C, low viscosity of 0.33 mm^2/s (at 20 °C) and a flash point of –18 °C, OME1 satisfies neither the hydraulic requirements with regard to high-pressure fuel injection nor the existing statutory safety provisions. Most research projects therefore focus on the synthesis possibilities and engine properties of the longer-chain ethers OME3-5.

OME3-5

Mixtures of the OME3-5 molecules have a boiling range that equates to a large extent to conventional diesel fuels. This opens up the prospect of being able to utilize the emission advantage of OME with fewer changes to the engine. Two properties of OME nevertheless prevent compatibility with the on-the-road vehicle population. The energy content of OME3-5 (H_i of OME3: 19.4 MJ/kg ; H_i of OME4: 18.7 MJ/kg; H_i of OME5: 18.1 MJ/kg) is less than half compared with diesel fuel (H_i of 43.0 MJ/kg), meaning that for the same engine power roughly 70 % higher volumetric injected fuel quantity is required. Furthermore, OME is not compatible with fluorocaoutchouc polymers, which are widely used as elastomer material in diesel fuel-injection equipment [36].

Engines with adapted materials and OME-specific calibration demonstrate significant advantages in untreated emissions, particularly low sooting. During combustion the formation of soot and NO_x is in reciprocal dependence of each other (trade-off). The application of high exhaust-gas recirculation rates permits a reduction of the NO_x emissions, which may be accompanied by a significant increase in soot particulates. This is not the case for OME because the formation of soot is greatly limited due to the carbon atoms isolated by oxygen bridges. Therefore both the concentration of NO_x and the number of soot particulates in the exhaust gas of OME can be greatly reduced [37]. The suitability of OME as a new fuel has also be verified in vehicle demonstrators [38].

To provide a basis for a future uniform quality for OME3-5, a tentative standard (DIN/TS 51699, [39]) is currently being developed in which the questions that are still open for example regarding thermo-oxidative stability, cold resistance, water content or limiting the content of OME2 and OME6 are addressed. Even a limit value for the total formaldehyde content must be introduced.

Unlike OME1, it is difficult to synthesize the higher oxymethylene ethers OME3-5. The equilibrium of starting materials, intermediates and end products arising from the reaction has a bearing on the chain length. During conversion the amount of formaldehyde that is formed must be controlled precisely, the water created removed, and released hydrogen recovered. For sustainability reasons the distillation outlay must also be kept as low as possible. The synthesis of OME3-5 is the subject of numerous research projects [40].

Table 6: Properties of liquid fuels and hydrocarbons

Material/medium	Density	Main constituents	Boiling temperature	Specific heat of vaporization	Net calorific value	Ignition temperature	Air requirement, theoretical	Ignition limit	
								Lower	Upper
	kg/*l*	% (m/m)	°C	kJ/kg	MJ/kg	°C	kg/kg	% (V/V) gas in air	
Gasoline,									
Regular	0.720...0.775	86 C, 14 H	25...210	380...500	41.2...41.9	≈ 300	14.8	≈ 0.6	≈ 8
Premium	0.720...0.775	86 C, 14 H	25...210	–	40.1...41.6	≈ 400	14.7	–	–
Aviation fuel	0.720	85 C, 15 H	40...180	–	43.5	≈ 500	–	≈ 0.7	≈ 8
Kerosene	0.77...0.83	87 C, 13 H	170...260	–	43	≈ 250	14.5	≈ 0.6	≈ 7.5
Diesel fuel	0.820...0.845	86 C, 14 H	180...360	≈ 250	42.9...43.1	≈ 250	14.5	≈ 0.6	≈ 7.5
Mineral oil (crude oil)	0.70...1.0	80...83 C, 10...14 H	25...360	222...352	39.8...46.1	≈ 220	–	≈ 0.6	≈ 6.5
Lignite tar oil	0.850...0.90	84 C, 11 H	200...360	–	40.2...41.9	–	13.5	–	–
Coaltar oil	1.0...1.10	89 C, 7 H	170...330	–	36.4...38.5	–	–	–	–
Pentane C_5H_{12}	0.63	83 C, 17 H	36	352	45.4	285	15.4	1.4	7.8
Hexane C_6H_{14}	0.66	84 C, 16 H	69	331	44.7	240	15.2	1.2	7.4
n-heptane C_7H_{16}	0.68	84 C, 16 H	98	310	44.4	220	15.2	1.1	6.7
iso-octane C_8H_{18}	0.69	84 C, 16 H	99	297	44.6	410	15.2	1	6
Benzene C_6H_6	0.88	92 C, 8 H	80	394	40.2	550	13.3	1.2	8
Toluene C_7H_8	0.87	91 C, 9 H	110	364	40.6	530	13.4	1.2	7
Xylene C_8H_{11}	0.88	91 C, 9 H	144	339	40.6	460	13.7	1	7.6
Ether $(C_2H_5)_2O$	0.72	64 C, 14 H, 22 O	35	377	34.3	170	7.7	1.7	36
Acetone $(CH_3)_2CO$	0.79	62 C, 10 H, 28 O	56	523	28.5	540	9.4	2.5	13
Ethanol C_2H_5OH	0.79	52 C, 13 H, 35 O	78	904	26.8	420	9	3.5	15
Methanol CH_3OH	0.79	38 C, 12 H, 50 O	65	1110	19.7	450	6.4	5.5	26
Dimethoxymethane (OME1)	0.86	47 C, 11 H, 42 O	42	376	22.7	235	7.2	2.2	19.9
Rape oil	0.92	78 C, 12 H, 10 O	–	–	38	≈ 300	12.4	–	–
Rapeseed methyl ester (biodiesel)	0.88	77 C, 12 H, 11 O	320…360	–	36.5	283	12.8	–	–

Viscosity at 20 °C in mm^2/s (= cSt): gasoline ≈ 0.6; ethanol ≈ 1.5; methanol ≈ 0.75.

Table 7: Properties of gaseous fuels and hydrocarbons

Material/medium		Density at 0 °C and 1.013 mbar	Main constituents	Boiling temperature at 1.013 mbar	Net calorific value	Ignition temperature	Air requirement, theoretical	Ignition limit	
								Lower	Upper
		kg/m³	% (m/m)	°C	MJ/kg	°C	kg/kg	% (V/V) gas in air	
Liquefied petroleum gas		2.25 [1]	90 C_3H_8, 10 C_4H_{10}	– 30	46.1	≈ 400	15.7	1.5	15
Municipal gas		0.56...0.61	8 H_2, 20 CH_4, 45 CO, 2 C_2H_6, 13 CO_2, 8 N_2	–210	26.3	≈ 560	7.6	4	40
Natural gas H (North Sea)		0.83	78 CH_4, 11 C_2H_6, 3 C_3H_8, 1 C_4H_{10}, 5 CO_2, 1 N_2	–162 (CH_4)	46.7	584	16.0	4.0	15.8
Natural gas H (Russia)		0.73	94 CH_4, 3 C_2H_6, 1 C_3H_8, 1 C_4H_{10}, 1 N_2	–162 (CH_4)	49.1	619	16.8	4.3	16.2
Natural gas L		0.83	75 CH_4, 5 C_2H_6, 1 C_3H_8, 4 CO_2, 14 N_2	–162 (CH_4)	40.7	≈ 600	14.0	4.6	16.0
Water gas		0.71	6 H_2, 71 CO, 14 CO_2, 8 N_2	–	14.8	≈ 600	3.8	6	72
Blast-furnace gas		1.28	30 CO, 14 CO_2, 56 N_2	–170	3.3	≈ 600	0.75	≈ 30	≈ 75
Sewage gas (manure gas)		–	40 CH_4, 60 CO_2	–	20.1	–	6.9	–	–
Hydrogen	H_2	0.090	100 H	–253	120.0	560	34.3	4	77
Carbon monoxide	CO	1.25	100 CO	–191	10.05	605	2.5	12.5	75
Methane	CH_4	0.72	75 C, 25 H	–162	50.0	650	17.2	5	15
Acetylene	C_2H_2	1.17	93 C, 7 H	– 81	48.2	305	13.3	1.5	80
Ethane	C_2H_6	1.36	80 C, 20 H	– 88	47.5	515	16.1	3	14
Ethene	C_2H_4	1.26	86 C, 14 H	–102	47.1	425	14.8	2.75	34
Propane	C_3H_8	2.0 [1]	82 C, 18 H	– 43	46.3	470	15.7	1.9	9.5
Propene	C_3H_6	1.92	86 C, 14 H	– 47	45.8	450	14.8	2	11
Butane	C_4H_{10}	2.7 [1]	83 C, 17 H	– 10; +1 [2]	45.6	365	15.5	1.5	8.5
Butene	C_4H_8	2.5	86 C, 14 H	– 5; +1 [2]	45.2	–	14.8	1.7	9
Dimethyl ether	C_2H_6O	2.05 [3]	52 C, 13 H, 35 O	– 25	28.8	235	9.0	3.4	18.6

[1] Density of liquefied gas 0.54 kg/*l*, density of liquid propane 0.51 kg/*l*, density of liquid butane 0.58 kg/*l*.
[2] First value for isobutane, second value for n-butane and n-butene.
[3] Density of liquefied dimethyl ether 0.667 kg/*l*.

References

[1] DIN EN 228:2017+ Corrigendum 2020; Automotive fuels – Unleaded petrol – Requirements and test methods.
[2] ASTM D4814-20, Standard Specification for Automotive Spark-Ignition Engine Fuel.
[3] DIN EN ISO 5164:2014, Petroleum products – Determination of knock characteristics of motor fuels – Research method (ISO 5164:2014).
[4] DIN EN ISO 5163:2014, Petroleum products – Determination of knock characteristics of motor and aviation fuels – Motor method (ISO 5163:2014).
[5] J. Dedl, B. Geringer, O. Budak et al.: Kraftstoffkennzahlen zur Beschreibung von Vorentflammung in Ottomotoren, MTZ 79, 76–81 (2018).
[6] DIN EN ISO 3405:2019, Petroleum and related products form natural or synthetic sources – Determination of distillation characteristics at atmospheric pressure (ISO 3405:2019).
[7] DIN EN 13016-1:2018, Liquid petroleum products – Vapour pressure – Part 1: Determination of air saturated vapour pressure (ASVP) and calculated dry vapour pressure equivalent (DVPE).
[8] DIN EN 15293:2018, Automotive fuels – Automotive ethanol (E85) – Requirements and test methods.
[9] ASTM D5798-20, Standard Specification for Ethanol Fuel Blends for Flexible-Fuel Automotive Spark-Ignition Engines.
[10] DIN EN 590:2017, Automotive fuels – Diesel – Requirements and test methods.
[11] DIN EN 14214:2019, Liquid petroleum products – Fatty acid methyl esters (FAME) for use in diesel engines and heating applications – Requirements and test methods.
[12] ASTM D975-20, Standard Specification for Diesel Fuel.
[13] ASTM D6751-20, Standard Specification for Biodiesel Fuel Blend Stock (B100) for Middle Distillate Fuels.
[14] DIN EN ISO 5165:2018, Petroleum products – Determination of the ignition quality of diesel fuels – Cetane engine method (ISO 5165:2017).
[15] DIN EN 15195:2015, Liquid petroleum products – Determination of ignition delay and derived cetane number (DCN) of middle distillate fuels by combustion in a constant volume chamber.
[16] DIN EN 16144:2012, Liquid petroleum products – Determination of ignition delay and derived cetane number (DCN) of middle distillate fuels – Fixed range injection period, constant volume combustion chamber method.
[17] DIN EN 16715:2015, Liquid petroleum products – Determination of ignition delay and derived cetane number (DCN) of middle distillate fuels – Ignition delay and combustion delay determination using a constant volume combustion chamber with direct fuel injection.
[18] DIN EN 17155:2018-09, Liquid petroleum products – Determination of indicated cetane number (ICN) of middle distillate fuels – Primary reference fuels calibration method using a constant volume combustion chamber.
[19] SS 155435:2016, Automotive fuels – Diesel fuel oil of environmental class 1 and 2 for high-speed diesel engines – Requirements and test methods.
[20] DIN EN ISO 12156-1:2019, Diesel fuel – Assessment of lubricity using the high-frequency reciprocating rig (HFRR) – Part 1: Test method (ISO 12156-1:2018).
[21] DIN EN 15751:2014, Automotive fuels – Fatty acid methyl ester (FAME) fuel and blends with diesel fuel – Determination of oxidation stability by accelerated oxidation method.

[22] FPRCEN/TR 17548:2020, Automotive fuels – Diesel fuel market issues – Abrasive particles investigation report.
[23] DIN EN 16709:2019, Automotive fuels – High FAME diesel fuel (B20 and B30) – Requirements and test methods.
[24] ASTM D7467-20, Standard Specification for Diesel Fuel Oil, Biodiesel Blend (B6 to B20).
[25] DIN EN 16734:2019, Automotive fuels – Automotive B10 diesel fuel – Requirements and test methods.
[26] L. König, J. Gaube: Fischer-Tropsch-Synthese, Neuere Untersuchung und Entwicklungen. Article in Chemie Ingenieur Technik, Volume 55/1, 1983.
[27] DIN EN 15940:2019, Automotive fuels – Paraffinic diesel fuels from synthesis or hydrotreatment – Requirements and test methods.
[28] K. Altfeld, P. Schley: Entwicklung der Erdgasbeschaffenheit in Europa. Gaswärme International, 58-63 (February 2012).
[29] DVGW, Gasbeschaffenheit, Technische Regel – Arbeitsblatt DVGW G260 (A), March 2013.
[30] DIN EN 16723-2:2017, Natural gas and biomethane for use in transport and biomethane for injection in the natural gas network – Part 2: Automotive fuels specification.
[31] K. Reif (Editor): Ottomotor-Management, 4th Ed., Verlag Springer Vieweg 2014.
[32] J. Förster: Diesel-Gasmotoren. In: K. Mollenhauer, H. Tschöke, R. Maier (Editors): Handbuch Dieselmotoren, 4th Ed., pp. 219–224, Verlag Springer Vieweg, 2018.
[33] J. Förster: Direkteinblasung bei Dieselmotoren. In: K. Mollenhauer, H. Tschöke, R. Maier (Editors): Handbuch Dieselmotoren, 4th Ed., pp. 241–247, Verlag Springer Vieweg, 2018.
[34] DIN EN 589:2019, Automotive fuels – LPG – Requirements and test methods.
[35] W. Maus (Editor): Zukünftige Kraftstoffe – Energiewende des Transports als ein weltweites Klimaziel, ATZ/MTZ-Fachbuch, Verlag Springer Vieweg 2019.
[36] M. Härtl, K. Gaukel, D. Pélerin et al.: Oxymethylenether als potenziell CO_2-neutraler Kraftstoff für saubere Dieselmotoren Teil 1: Motorenuntersuchungen, MTZ 78, 52–59 (2017).
[37] D. Pélerin, M. Härtl, E. Jacob, G. Wachtmeister: Potentials to simplify the engine system using the alternative diesel fuels oxymethylene ether OME1 and OME3–6 on a heavy-duty engine, Fuel 259, 116231 (2020).
[38] M. Münz, A. Mokros, D. Töpfer et al.: OME – Partikelbewertung unter Realfahrbedingungen, MTZ 79, 16–21 (2018).
[39] DIN/TS 51699, Fuels – Polyoxymethylene dimethyl ethers (OME) – Requirements and test methods.
[40] J. Burger, H. Hasse: Processes for the production of OME fuels, 7. International Engine Congress, Baden-Baden, 2020.

Urea/water solution

Application
A urea/water solution is used to reduce nitrogen oxides (NO_x) in the catalytic exhaust-gas treatment of diesel engines with the aid of SCR systems (see SCR system). To this end, this solution is not used as a fuel additive, but metered directly into the exhaust-system branch.

Designations
The solution used under the protected trade name AdBlue® consists of 32.5 % technically pure urea and 67.5 % demineralized water. The chemical total formula is: CH_4N_2O. Further designations are:
– carbamide in aqueous solution,
– Aqueous Urea Solution 32.5 % (AUS 32),
– NO_x reduction agent in acc. with DIN 70070 [1] or ISO 22241 [2] Part 1 to Part 4,
– urea/water solution,
– Diesel Exhaust Fluid (DEF),
– Agente Reductor Líquido de Óxido de Nitrogénio Automotivo (ARLA 32).

Properties
High temperatures and direct sunlight are to be avoided due to decomposition by heat. When AdBlue is heated above 25 °C, an ammonia solution is created as an aqueous solution or as an aerosol of ammonia as a hazard-determining substance.

AdBlue attacks base metals. Copper, alloys containing copper and unalloyed and galvanized steels, aluminum and glass are not suitable for contact with a urea/water solution.

AdBlue has a strong tendency to crystallization. White crystals form when the water that it contains evaporates. The growth of crystals is encouraged by increased air moisture. AdBlue has high penetration ability and a strong dispersion capacity. Lines/pipes can be blocked by crystallization. Blockages can be dissolved by warm water. It is essential to look out for the growth of crystals along electrical wiring/leads. Electric contacts can be jumpered by the high conductivity. Leads containing copper are corroded.

The product is not classified as hazardous and does not have to be specially labeled. The substance is not irritant, but after repeated or prolonged skin contact can cause degreasing of the skin and dermatitis. Experience in practice has shown reddened skin after a few minutes in contact with the medium.

With water hazard class 1, AdBlue is classified as slightly hazardous to water and may not be released into the environment.

At ambient temperatures below –10 °C containers, lines and equipment must be fitted with cold insulation and heating. Frozen AdBlue incurs no damage and can therefore be reused once it has thawed.

Storing AdBlue at temperatures below 0 °C is to be avoided due to the risk of freezing and the associated volumetric expansion. There is the danger of the container splitting. A leak may only become visible after thawing.

References
[1] DIN 70070: Diesel engines – NO_x reduction agent AUS 32 – Quality requirements.
[2] ISO 22241: Diesel engines – NO_x reduction agent AUS 32 – Quality requirements.

Table 1: Properties of AdBlue

Delivery form	Aqueous solution
Appearance	Clear and colorless
Odor	Odorless up to slight odor of ammonia
Solubility	Mixable with water in any ratio, poorly soluble in aliphatic hydrocarbons
Congealing point	–11.5 °C
Reaction	Weakly alkaline
Electric conductivity	Liquid and crystals good electric conductivity

Refrigerants for air-conditioning systems

The core component of the air-conditioning system in motor vehicles is as a rule a cold vapor-compression refrigerating system (see Passenger-compartment climate control). The underlying thermodynamic cyclic process – the Carnot cycle – is approximated with this system (see Thermodynamics). The two isentropic changes of state (according to Carnot) are approximated by a air compressor and a throttle element and the isotherms are approximated by evaporating and condensing refrigerant. Operating fluids for which the phase changes occur at tenable pressures are required especially for evaporation and condensation.

Table 1 shows the property values of the refrigerants R134a and R1234yf. The global warming potential (GWP) is a measure of the relative contribution of a greenhouse gas to the greenhouse effect. The reference forms the CO_2 with the global warming potential of one. The emission of 1 kg R134a would accordingly have a greenhouse effect comparable with the emission of 1300 kg CO_2.

Table 1: Property values of the refrigerants R134a and R1234yf

Property	R134a	R1234yf
Boiling point	–26 °C	–29 °C
Vapor pressure at 25 °C	6.56 bar	6.64 bar
Vapor pressure at 80 °C	25.97 bar	24.38 bar
Vapor density	32.4 kg/m³	37.6 kg/m³
GWP	approx. 1300	approx. 4

Development up to today

Refrigerant R12

Initially, combustible substances (e.g., methyl chloride CH_3Cl, sulfur dioxide SO_2 or diethyl ether $C_4H_{10}O$) were used in refrigerant systems. In around 1930 the American engineer Thomas Midgley developed chlorofluorocarbons (CFCs). The most important substance from this family for motor vehicles was R12 (CCl_2F_2). This substance accepted as non-poisonous and non-combustible was used in practically all passenger-car air-conditioning systems and was additionally available for a great many other applications.

In 1974 the chemists Molina and Rowland put forward the hypothesis that substances containing chlorine damage the ozone layer that protects us [1]. Subsequent observations of the ozone hole over Antarctica bore out this hypothesis. The evaluation of precise measurements finally provided certainty. This resulted on 16 September 1987 (in Montreal) in the signing of the Montreal Protocol, in which the resolution was passed to renounce chlorine chemistry.

Refrigerant R134a

From around 2000 onwards, the refrigerant R134a ($C_2H_2F_4$) was used as a substitute substance. But heightened sensitivity to all environmental impacts has also led to the limitation or restriction on the use of R134a. According to an EU Directive [2], air-conditioning systems designed to contain fluorinated greenhouse gases with a global warming potential of over 150 are to be prohibited in motor vehicles. R134a has a global warming potential of around 1300 and therefore falls below this Directive. Since January 2011 new vehicle types have no longer been permitted to be equipped with R134a and since January 2017 all new cars have been required to be equipped with a refrigerant that has a global warming potential of less than 150. Also applicable aside from this Directive is the EC Regulation [3], the main purpose of which is the avoidance or reduction of the emission of fluorinated greenhouse gases.

Refrigerant R1234yf
At the moment the only new refrigerant available that is produced in sufficient quantities is R1234yf ($C_3H_2F_4$) with a global warming potential of 4. Since 2017 it has, apart from a very few exceptions, been used by all manufacturers in all passenger-car air-conditioning systems. The very low global warming potential chiefly results from the fact that this substance has a very short service life in the atmosphere. R1234yf is however not undisputed. It is more easily flammable than R134a and in the event of a fire caustic and toxic reaction products are created. Therefore automobile manufacturers take additional safety precautions to ensure the customarily high safety of occupants.

Refrigerant CO_2
Air-conditioning systems with CO_2 as refrigerant are used in some upper-category passenger cars. CO_2 has by definition a global warming potential of 1 and is thus naturally predestined to be used. Unfortunately a CO_2 air-conditioning system cannot be operated with the same process and the same components. The process works on the high-pressure side in the supercritical range at very high pressures (100 to 150 bar). Therefore all the components must be newly developed and dimensioned for these unusually high pressures. Since such new developments affect not only development, production and sales but also to a significant extent service, a comprehensive market launch on 1st January 2017 could not be achieved.

An additional risk also arises in CO_2 systems. Because of the extremely high pressures a considerable amount of energy (pressure times volume) is stored in the system. This energy could for example in an accident be additionally released and cause further damage. Therefore enhanced safety precautions must also be taken for CO_2 air-conditioning systems.

Alternative possibilities

It is safe to assume that R1234yf will not be the last refrigerant used. One possibility is to build on the experiences with the CO_2 air-conditioning systems and expand this technology as a long-term solution. If this does not prove successful, natural substances which are really neutral in their impact on the environment can be considered. There have already been studies in this connection at the time of the discussions regarding renunciation of chlorine chemistry when scientists were locking for substitute substances for refrigerants containing chlorine [4]. The most natural substances would be water and air.

Water as refrigerant
In principle, water could be used in a cold vapor-compression refrigerating system, but evaporation and condensation would take place at extremely low pressures. In other words, the tiniest leaks would raise the pressure and the system function would no longer be guaranteed. The function has been verified in the laboratory. However, this would not be easily feasible at the current time.

Air as refrigerant
Air could be used in systems which operate according to the Joule process. Such systems are installed in virtually all commercial aircraft. However, the design of systems with much smaller output figures – i.e. for passenger-car applications – is more difficult and at the least entails diminished effectiveness. The process has been used with a certain degree of success for climate control in rail vehicles, for example in the ICE 1. The system output here is on a comparable scale to aircraft, but conventional R134a systems have a significantly better effectiveness.

At the moment the German Federal Environment Agency is supporting a research project in which systems with natural refrigerants are to be reassessed [5]. Since the Joule process is a pure gas process, it is inferior to the Carnot process in terms of effectiveness. It is used economically in aircraft due to the fact that some positive supplementary effects

are utilized. The air-conditioning system here has the additional task of pressurizing the cabin. The system weighs less than half as much as a comparable cold vapor-compression system and the drive can be effectively realized with compressed air from the engines.

Combustible substances as alternative refrigerants

Combustible substances could be further alternatives. In refrigerators R12 for example has been superseded by isobutane. In motor vehicles, however, the situation is a little more difficult in that as the result of a leak for example in the evaporator combustible substances would inevitably get into the passenger cell.

References

[1]: M.J. Molina; F.S. Rowland: Stratospheric sink for chlorofluoromethanes: chlorine atom catalysed destruction of ozone. Nature, No. 249 pp 810–812, June 28, 1974.

[2] Directive 2006/40/EC of the European Parliament and of the Council of 17 May 2006 relating to emissions from air-conditioning systems in motor vehicles and amending Council Directive 70/156/EEC.

[3] Regulation (EC) No. 842/2006 of the European Parliament and of the Council of 17 May 2006 on certain fluorinated greenhouse gases.

[4] H. Kruse: Alternative Kälteprozesse unter Umweltschutzgesichtspunkten – DKV status report of the Deutscher Kälte- und Klimatechnischer Verein No. 3 (1989).

[5] Umweltbundesamt: Erprobung, Messung und Bewertung von Systemen mit natürlichen Kältemitteln zum nachhaltigen Kühlen und Heizen von öffentlichen Verkehrsmitteln – Ersatz fluorierter Treibhausgase. Duration 07 May 2015 – 31 March 2018.

Drivetrain with internal-combustion engine

Drivetrain elements

The drivetrain is in automotive engineering assigned all the components which generate power for drive purposes and transmit that power to the road. The conventional drivetrain of a motor vehicle includes the internal-combustion engine, the transmission, the propeller shaft, the final-drive differential, and the drive shafts. The transmission contains a vibration damper, a power take-up element, and a step-up gear unit with gear-shift elements. Figure 1 shows the arrangement of these components for a standard drive (rear-wheel drive). It also shows the components for an all-wheel drive.

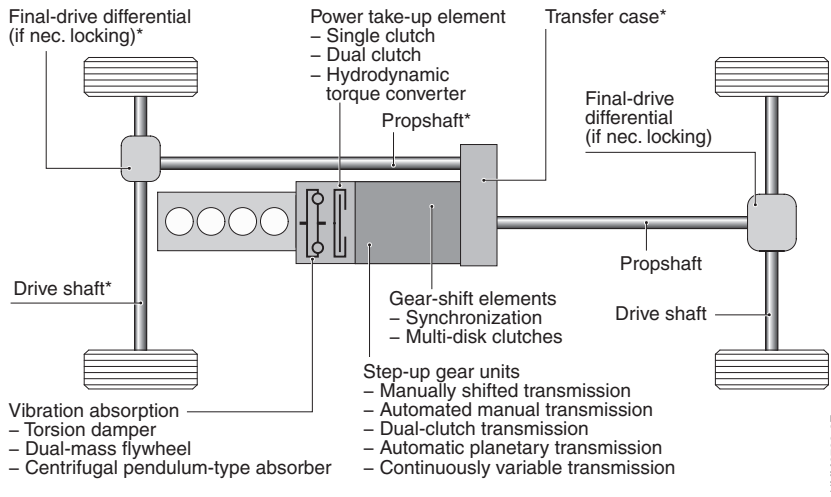
Hybrid drives feature an electric machine in addition to the internal-combustion engine. The components in the drivetrain that concern the internal-combustion engine are comparable with the conventional drive.

Functions of the drivetrain

The function of a drivetrain is to adapt the power outputs and torques made available by the engine to the respective driving situation and to the driver's commands. The transmission converts the torque into the tractive force for the wheels required to overcome the running resistances. For practical driving it must be possible to set different gear ratios in the step-up gear unit.

The internal-combustion engine converts the energy of the fuels supplied to it into mechanical energy, operates between a minimum and a maximum rotational speed in the process, and delivers a limited torque in this range. The rotational speed of internal-combustion engines is so high that their crankshafts are unable to drive the drive wheels directly. The step-up gear unit and the final-drive differential convert the engine speed to a range that produces wheel speeds suitable for driving. The overall gear ratio i_A (ratio of engine speed to wheel speed) is less than 1; the torque output to the wheels in relation to the engine

Figure 1: Drivetrain elements for a standard drive (rear-wheel drive)
Components marked * are optionally provided for an all-wheel drive.



torque is increased in the same ratio. The transmission acts as a torque converter.

Unlike steam engines or electric machines, the internal-combustion engine cannot generate torque when it is at a standstill. Moreover, fuel consumption is very much dependent on the requested torque at the applied engine speed. Consequently, a drivetrain has further functions to fulfill:

– ensuring idle operation of the internal-combustion engine when the vehicle is at a standstill,
– facilitating driving off from a standstill,
– isolating the drivetrain from the rotational irregularities of the internal-combustion engine,
– reversing the direction of rotation of the drive wheels where necessary (reverse gear).

Tractive-force curves

Requirements

The requirements that a motor vehicle is to fulfill with regard to maximum climbing capability and acceleration on the one hand and top speed on the other hand define the necessary tractive force on the wheels and hence the demand on the transmission with regard to its spacing between the lowest and highest gears. The final-drive differential assumes some of the stepping-up duties so that the speed ratio in the step-up gear unit does not get too big for the different gears. The overall ratio i_A of the drivetrain is the product of the gear ratio i_n of the step-up gear unit in the selected gear and the constant gear ratio i_B of the final-drive differential.

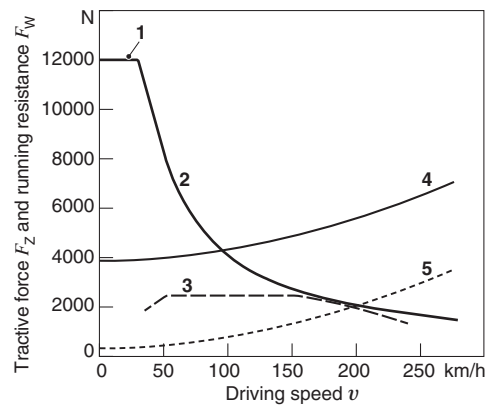
Tractive-force hyperbola

The tractive-force hyperbola (Figure 2) shows as a function of the driving speed v and the drivetrain efficiency η_A the maximum available tractive force F_Z for the event that the internal-combustion engine could output its maximum power P_{max} over its entire speed range:

$$F_Z = \frac{P_{max}}{v}\,\eta_A.$$

Figure 2: Tractive force as a function of speed for an internal-combustion engine without step-up gear unit (i.e., gear ratio of step-up gear unit equals 1)

1 Adhesion limit F_{Zmax},
2 Tractive-force hyperbola F_Z,
3 Tractive force F_Z for engine without step-up gear unit,
4 Characteristic running-resistance curve at gradient angle $\alpha = 10°$,
5 Characteristic running-resistance curve at gradient angle $\alpha = 0°$.



In the lower speed range the tractive force is limited not by the tractive-force hyperbola but by the adhesion limit F_{Zmax} between tire and road which is obtained from the product of the normal force F_N of the driven vehicle wheels and the coefficient of friction μ of the roadway:

$$F_{Zmax} = F_N \mu .$$

Influence of running resistances
The running resistance F_W of the vehicle without acceleration is made up of rolling resistance F_R, aerodynamic drag F_L, and climbing resistance F_S:

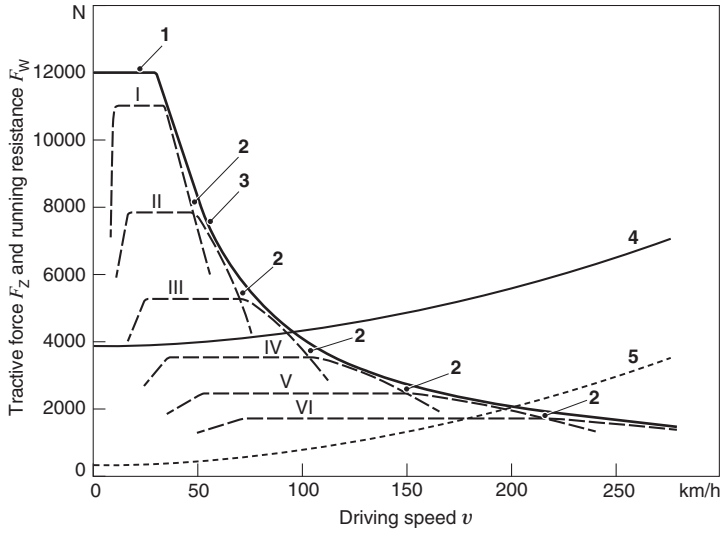
$$F_W = F_R + F_L + F_S .$$

The factors that influence the rolling resistance are the vehicle mass m, the gravitational acceleration g, the coefficient of friction μ, and the gradient angle α of the roadway. The aerodynamic drag contains the drag coefficient c_W, the frontal area of the vehicle A, the density ρ of the air, and the driving speed v. The factors that influence the climbing resistance are the vehicle mass m, the gravitational acceleration g, and the gradient angle α of the roadway. Consequently:

$$F_W = m g \mu \cos\alpha + 0.5 c_W A \rho v^2 + m g \sin\alpha .$$

In Figure 2 it becomes clear that an internal-combustion engine without a step-up gear unit would only operate in a very limited driving range. Here the tractive force F_Z of the engine only contains the transmission ratio of the differential so that the effect of torque conversion in the step-up gear unit can be clearly explained in the following.

Figure 3: Tractive force as a function of speed for an engine with six gear steps (v_{max} in fifth gear, plus an overdrive gear)
1 Adhesion limit F_{Zmax},
2 Non-utilizable driving range,
3 Tractive-force hyperbola F_Z,
4 Characteristic running-resistance curve at gradient angle α = 10°,
5 Characteristic running-resistance curve at gradient angle α = 0°.
I…VI Tractive-force curves for gears 1 through 6.



Adaptation by step-up gear unit

Manual transmission

A manual transmission with defined gear ratios of the selectable gears (manually shifted transmission, dual-clutch transmission, automatic planetary transmission) effects an approximation of the maximum engine torque – which the internal-combustion engine develops between its minimum and maximum speeds – to the tractive-force hyperbola (Figure 3). The highest and the lowest gear ratios i defined the spacing φ_G of the transmission:

$$\varphi_G = \frac{i_{max}}{i_{min}} .$$

The tractive-force curves for the individual gears increase in the range above idle speed with increasing engine torque and drop off in the high speed range in response to decreasing engine torque.

Table 1: Quantities and units

Quantity		Unit
c_W	Drag coefficient	–
g	Gravitational acceleration	m/s²
i_A	Overall ratio of drivetrain	–
i_n	Transmission ratio of a gear	–
i_B	Gear ratio of final-drive differential	–
m	Vehicle mass	kg
n	Engine speed	min^{-1}
v	Driving speed	m/s
A	Frontal area of vehicle	m²
F_Z	Tractive force at wheels	n
F_N	Normal force of vehicle	n
F_W	Running resistance	n
F_R	Rolling resistance	n
F_L	Aerodynamic drag	n
F_S	Climbing resistance	n
M	Torque	N m
P	Engine power	kW
α	Gradient angle of roadway	degree
φ	Gear step	–
φ_G	Transmission spacing	–
η	Efficiency	–
μ	Coefficient of friction of roadway	–
ρ	Density of the air	kg/m³

Both geometrical and progressive gear stepping can be implemented with the number of gears and the individual transmission ratios. The gear step φ between the gears is defined as the ratio of transmission i of two neighboring gears:

$$\varphi = \frac{i_{n-1}}{i_n} .$$

In the case of geometrical gear stepping the gear step is always the same. This is used in truck transmissions with high numbers of gears. A truck requires fine stepping in the entire driving range since its engine power is relatively low with regard to its high mass.

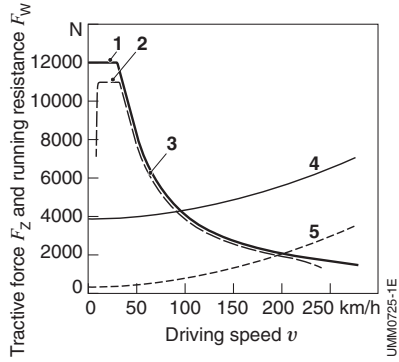
Progressive gear stepping is used in passenger-car transmissions. This means that the gear step becomes increasingly smaller to the higher gears. Consequently, the non-utilizable driving ranges (see Figure 3) are kept small in the car's main driving range although the number of gears is much smaller than in a truck. To achieve good driveability the tractive-force curves of the individual gears should intersect in each speed range (see Figure 3).

The point of intersection of the characteristic running-resistance curve at a 0° gradient angle with the tractive-force hyperbola defines the vehicle's top speed v_{max}. The ratio of the gear in which the top speed is to be reached should be selected in such a way that the engine outputs its maximum power at this point (fifth gear in Figure 3).

To reduce fuel consumption there are one or more gears with an overdrive ratio (sixth gear in Figure 3). In this gear the top speed cannot be reached (in Figure 3 a maximum speed of approximately 180 km/h on a 0° gradient is possible), but reducing the engine speed lowers the fuel consumption in a wide road-speed range.

Continuously variable transmission
A continuously variable transmission (CVT) can make any number of transmission ratios available. In this way, the maximum engine tractive force is introduced to the tractive-force hyperbola (Figure 4). On account of their design, continuously variable transmissions have a limited transmission spacing and a lower transmission efficiency. In Figure 4 the highest ratio corresponds roughly to the ratio of 5th gear of the multi-step transmission in Figure 3. Consequently, above all in the upper speed ranges, there is no possibility of reducing the engine speed and with it fuel consumption. This downturn in efficiency of the continuously variable transmission is however overcompensated for by higher overall-system efficiency (see Continuously variable transmissions).

Figure 4: Tractive force as a function of speed for an engine with continuously variable transmission
1 Adhesion limit F_{Zmax},
2 Tractive force F_Z for continuously variable transmission,
3 Tractive-force hyperbola F_Z,
4 Characteristic running-resistance curve at $\alpha = 10°$,
5 Characteristic running-resistance curve at $\alpha = 0°$.



Drive types

The arrangement of the engine and the step-up gear unit with vibration absorber and power take-up element depends very much on the vehicle category.

Standard drive
In a standard drive the engine with the step-up gear unit is arranged at the front in the front-axle area and the final-drive differential is arranged on the rear axle. The wheels on the rear axle are driven via the propeller shaft, the final-drive differential, and the drive shafts (rear-wheel drive, Figure 5a). The engine is longitudinally installed so that there are few limitations on engine length (e.g., 6-cylinder in-line engine) and thus on engine power output.

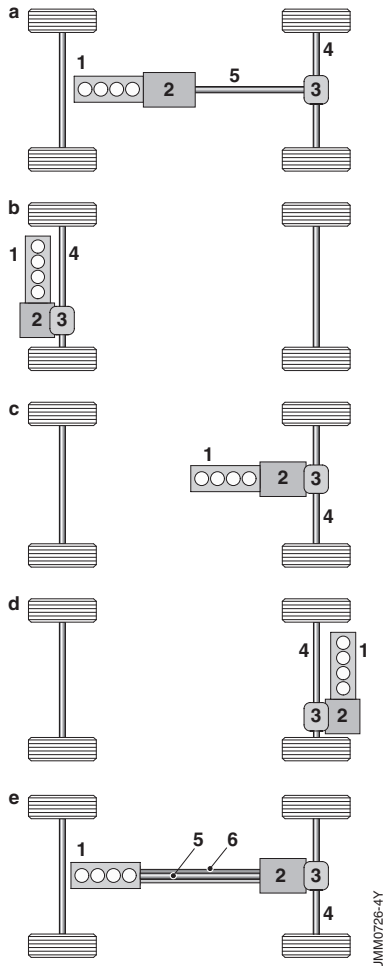
Until around 1930 this drive type was customary in virtually all vehicles; the designation standard drive dates back to that era. Since then, this drive type has featured almost exclusively in the upper mid-size and luxury categories.

Front-wheel drive
The front-wheel drive has in the meantime gained acceptance in vehicles ranging from the subcompact-size to the lower mid-size category and now features in more than 75 % of worldwide vehicle production. The engine and the step-up gear unit are generally arranged transversally to the direction of travel in the engine compartment (Figure 5b), but this entails limitations on engine size. Longitudinally installed engines are the exception. The steered front wheels are driven via the final-drive differential and the drive shafts. The final-drive differential is integrated in the step-up gear unit.

All-wheel drive
The all-wheel drive of many off-road vehicles is a variant of the standard drive with a transfer case situated in the direction of power flow after the step-up gear unit which transfers the transmission output torque to the front and rear axles (see Figure 1).

All-wheel drive is also possible in conjunction with a rear engine. All-wheel drive with a mid engine is difficult owing to the limited space available, therefore only a few vehicles feature this drive type. All-wheel drive with a front engine can be achieved

Figure 5: Arrangement of engine and transmission
a) Standard drive (longitudinal arrangement of engine),
b) Front-wheel drive (transversal arrangement of engine),
c) Mid-engine drive,
d) Rear-engine drive,
e) Transaxle drive.
1 Engine,
2 Step-up gear unit with vibration absorber and power take-up element,
3 Final-drive differential,
4 Drive shaft,
5 Propeller shaft,
6 Tube.



with a transfer case mounted on the transmission.

Mid-engine drive
The engine is installed in front of the powered rear axle. There are vehicles with transversally mounted engines, but also some with longitudinally installed engines. The step-up gear unit and the final-drive differential are situated in the rear-axle area (Figure 5c).

Mid engines with rear-wheel drive are installed almost exclusively in sports vehicles.

Rear-engine drive
The engine, the step-up gear unit and the final-drive differential are situated in the area of the powered rear axle (Figure 5d). Rear engines with rear-wheel drive are in the meantime installed almost exclusively in sports vehicles.

The engine can be longitudinally or transversally installed at the rear.

Transaxle drive
In the case of the transaxle drive, the engine is installed at the front and the step-up gear unit and the final-drive differential are mounted on the rear axle (Figure 5e). The engine and the power take-up element with step-up gear unit are connected by a tube and together form a rigid drive unit. The tube contains the non-articulated drive shaft, via which the transmission is driven. The drive shaft rotates with the engine speed.

An advantage of this arrangement of transmission and engine is uniform axle-load distribution. This drive is used in sports cars.

The inverted transaxle drive is also possible with a mid engine; the step-up gear unit and the final-drive differential are mounted here on the front axle.

Underfloor-engine drive
An underfloor layout is possible in commercial vehicles. The engine is situated between the axles; in a truck it is arranged on the frame. Uniform axle-load distribution results in good traction even when the vehicle is empty. But, because the engine is difficult to access, this concept has not caught on.

Power take-up elements and vibration isolation

The internal-combustion engine exhibits a minimum speed. To drive off from a vehicle standstill, the rotational-speed gap between the lowest engine operating speed and the stationary transmission input shaft must be closed by a speed converter [1]. In addition, during driving the gearshifts in manual transmissions cannot happen under load. The functions of this power take-up element are therefore:
– interrupting the power flow between the engine and the step-up gear unit,
– soft and smooth starting,
– transmitting the engine torque at different engine speeds to the step-up gear unit.

The vibration-isolation components reduce torsional vibrations at the crankshaft and thereby reduce noise and wear in the drivetrain.

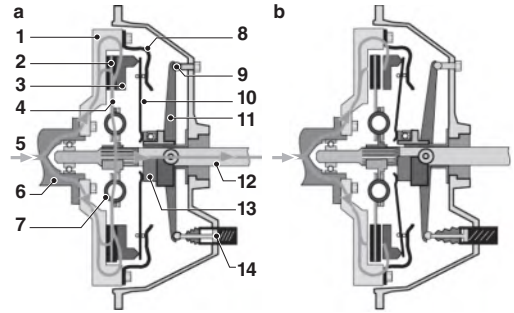
Single clutch

The clutch is the link between the engine and the step-up gear unit, transmitting the engine torque through adherence by frictional forces to the input shaft of the step-up gear unit.

Single-disk clutch
The single-disk clutch installed in most manually shifted transmissions consists of the clutch disk fitted with a clutch lining (Figure 2), which in the engaged state is pressed by a pressure pate pretensioned by the disk spring (diaphragm spring) against the engine flywheel (Figure 1a). A frictional connection is established between the engine flywheel and the clutch disk. The clutch disk is connected with torsional strength via the hub to the transmission input shaft.

The clutch is actuated in a manually shifted transmission via the clutch pedal. Pedal-force transmission in passenger cars is increasingly being performed by hydraulic systems and increasingly less so via control cable. Without an external operating force the diaphragm-spring force keeps the clutch closed. To open the clutch, it is necessary to depress the clutch pedal. The hydraulically actuated

Figure 1: Actuation of a single-disk clutch
a) Closed clutch,
b) Open clutch.
1 Engine flywheel,
2 Clutch disk with clutch linings,
3 Pressure plate with support for diaphragm spring,
4 Drive plate,
5 From engine,
6 Crankshaft,
7 Torsion damper,
8 Pressure-plate housing with support for diaphragm spring,
9 Pivot point for release fork,
10 Diaphragm spring with release tongues,
11 Release fork,
12 Transmission input shaft,
13 Release bearing,
14 Slave cylinder.

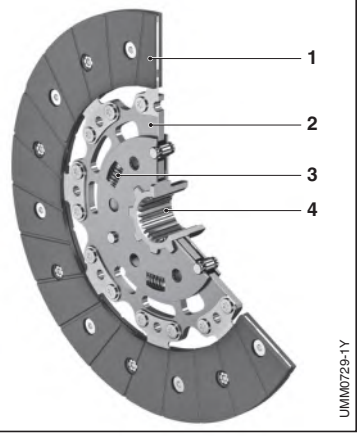


slave cylinder presses the release fork forwards (Figure 1b); this in turn pushes the release bearing and with it the inner part of the diaphragm spring forwards in the direction of the clutch disk. The outer part of the diaphragm spring moves in the opposite direction and in so doing pulls the pressure plate away from the clutch disk. In this way, as the pressure on the clutch pedal increases, the transmitted torque decreases until the clutch is fully opened.

When during driving off the force is slowly withdrawn via the clutch pedal, the diaphragm spring relaxes and increasingly presses the pressure plate against the engine flywheel. The transmitted torque increases and the vehicle drives off. When speed synchronization between the engine and the transmission has taken place, the drive-off process is completed and the clutch is fully engaged.

Torsion-damped clutch disks with torsional-vibration springs on the inside (Figure 2) damp engine vibrations and protect the transmission. They also serve to ensure a smooth clutch response when driving off. Single-disk clutches are used in both passenger cars and commercial vehicles.

Figure 2: Clutch disk with clutch lining
1 Clutch lining,
2 Clutch-disk drive plate,
3 Torsional-vibration springs (torsion damper),
4 Hub, to transmission.



Double-disk clutch

In rare cases there are versions for higher torques with two clutch disks (Figure 3). This version would be chosen if the single-disk clutch were too large in diameter or the contact pressure were too great to transmit the required torque.

Both clutch disks act on only one transmission input shaft. For this reason, the double-disk clutch too belongs to the category of single clutches.

Design types

Single and dual clutches can take the form of both a dry clutch and an oil-cooled “wet” variant.

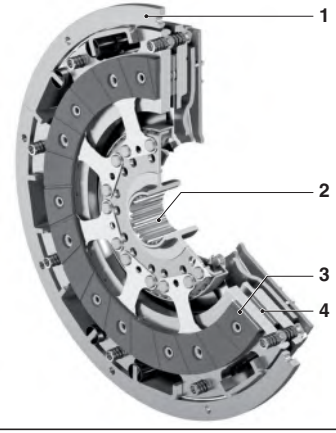
Dry clutch

Clutch linings for dry clutches consist for example of organic materials which are bound by resins. The materials used are similar to those of brake pads/linings. The aim is to achieve a high coefficient of friction for the clutches so as to keep the operating force and the diameter of the clutch small.

Wet clutch

Multi-disk clutches with paper-like linings are used for wet clutches. These clutches are similar to the power-shift elements in

Figure 3: Double-disk clutch
1 From driving engine,
2 Hub, to transmission,
3 Clutch disk 1 with lining,
4 Clutch disk 2 with lining.



automatic transmissions (see Gear-shift elements). The function of the paper-like lining is obtained by oil wetting. More heat can be dissipated via an oil flow through grooves in the clutch disks from the clutch than is the case with dry linings, where cooling is effected by thermal masses (flywheel, pressure plate) and front there via air convection.

Control devices

Hydraulic clutch actuation

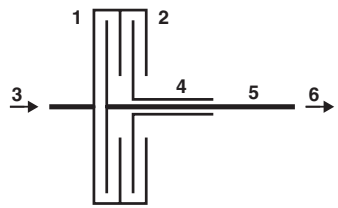
The master cylinder of the hydraulic control device is mechanically connected to the clutch pedal. The hydraulic fluid is withdrawn from the expansion tank for brake fluid. When the clutch pedal is pressed, pressure is built up in the clutch on the slave cylinder connected via a hose and the clutch opens as a result.

Clutch actuation for commercial vehicles

Large forces on the release bearing are required to actuate commercial-vehicle clutches. For light commercial vehicles compressed-air-boosted hydraulic systems are used to reduce the foot force on the clutch pedal. For heavy commercial vehicles electromechanical or electropneumatic systems are used [2].

Figure 4: Schematic diagram of a dual clutch

1 Clutch 1,
2 Clutch 2,
3 From vibration isolation, 4 Hollow shaft,
5 Solid shaft, 6 To transmission.



Dual clutch

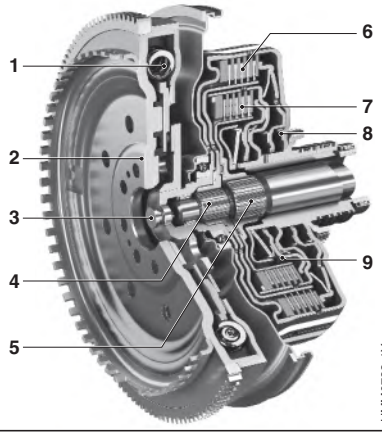
The dual clutch (Figure 4 and Figure 5) is characterized by the fact that two independently controllable clutches transmit the engine power to two transmission input shafts (a solid shaft and a hollow shaft). In most cases the dual clutch is combined with a dual-clutch transmission (see Dual-clutch transmission).

Figure 5 shows a dual clutch with oil-cooled multi-disk clutches. These are very similar clutch elements as are used as power-shift elements in automatic planetary transmissions. Actuation is effected by co-rotating pistons, which generate an axial force by means of hydraulic pressure. When the disks are pressed together with increasing axial force, the torque that can be transmitted via the clutch increases (see Multi-disk clutch).

Alternatively, the disks can also be pressed together via release bearings.

Figure 5: Structural design of a dual clutch

1 Torsion damper, 2 From driving engine,
3 From vibration isolation,
4 Solid shaft (connected to clutch 1),
5 Hollow shaft (connected to clutch 2),
6 Clutch 1,
7 Clutch 2,
8 Piston for clutch 1,
9 Piston for clutch 2.



Hydrodynamic torque converter

Design

The hydrodynamic torque converter (Figures 6 and 7) comprises the following components:

– impeller (connected to the internal-combustion engine),
– turbine (connected to the transmission input shaft),
– stator (connected via a one-way clutch to the transmission housing),
– lockup clutch, and
– integrated torsion damper.

Operating principle

The torque converter is filled with hydraulic oil. A gear pump delivers an oil pressure of 2 to 5 bar to prevent the oil from foaming and vapor pockets. The kinetic energy introduced by the engine is converted in the impeller into hydraulic energy. The faster the impeller driven by the engine rotates at speed n_P, the more oil is delivered by the impeller as the result of centrifugal force from the inside close to the rotational axis radially outwards. This oil flow strikes vanes

Figure 7: Structural design of hydrodynamic torque converter

1 Lockup clutch,
2 Torsion damper,
3 From driving engine,
4 To transmission input shaft,
5 Connection to transmission housing,
6 Impeller,
7 Turbine,
8 Stator,
9 One-way clutch.

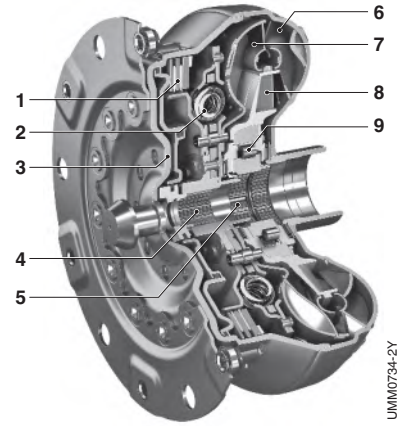
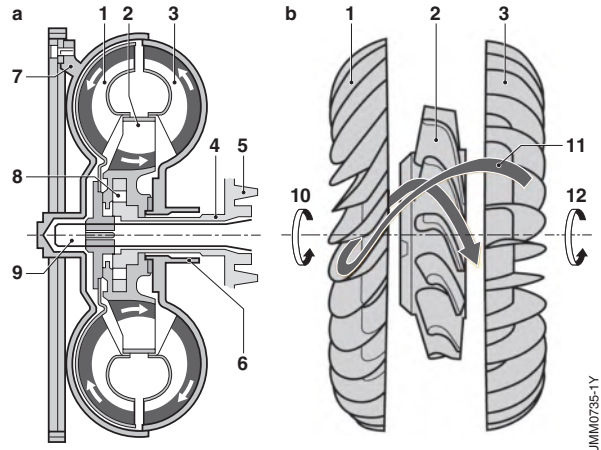


Figure 6: Hydrodynamic torque converter

a) Schematic diagram in cross-section,
b) Oil-flow course.
The turbine and the impeller rotate in such a way that the visible front moves downward. The stator, which would like to move in the conversion range in the opposite direction, is supported via a one-way clutch on the transmission housing.
The round arrows in the figure on the left symbolize the direction of the oil flow.

1 Turbine,
2 Stator,
3 Impeller,
4 Hollow shaft for stator support,
5 Housing (fixed),
6 Hollow impeller shaft for driving the transmission oil pump,
7 Converter cover connected to impeller,
8 Stator one-way clutch,
9 Turbine shaft (transmission inlet),
10 Direction of turbine rotation,
11 Oil flow direction,
12 Direction of impeller rotation.



in the neighboring turbine. The deflection of the oil flow to these vanes induces a torque which acts on the transmission input shaft. Because the conversion takes place using the hydrodynamic inertia of the oil, the hydrodynamic torque converter operates without wear.

The torque at the impeller is called pump torque M_P, that at the turbine is called turbine torque M_T.

Converter increase

The oil flow is delivered in the turbine from the outside radially inwards, where it strikes the stator (Figure 6). The stator features vanes which have a curvature of 90° and are supported by means of a one-way clutch in one direction of rotation on the transmission housing. While the turbine (and hence the transmission inlet) is stationary or only rotates slowly at speed n_T, the stator vanes induce a retention of oil which results in a reaction-torque increase at the turbine. With the oil flow being supported by means of the stator with stator torque M_L on the transmission housing, it is possible that a higher torque is transmitted to the transmission input shaft than the internal-combustion engine delivers to the converter. This is referred to as a converter increase. The following torque equilibrium applies to the torque converter:

$$M_P + M_T + M_L = 0$$

With regard to the torque converter the pump torque is an input torque and hence the preceding sign is positive, while the turbine torque is an output torque and hence the preceding sign is negative.

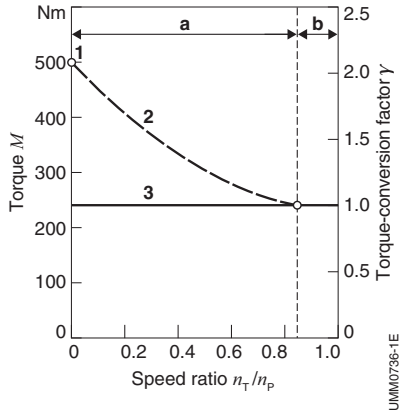
The deflection of the oil flow in the stator is greatest when the speed difference between the impeller and the turbine is greatest – i.e., when driving off. This is then followed by the greatest torque increase. As the speed of the turbine (i.e. the transmission input shaft) increases, the angle of impact of the oil flow on the stator vanes changes. The oil retention and hence the torque supported on the transmission housing are reduced. The converter increase decreases. The characteristic can be seen in Figure 8 in the torque-conversion factor γ:

$$\gamma = -\frac{M_T}{M_P}.$$

At the stall point when the turbine is stationary $1.7 < \gamma < 2.4$ usually applies.

As long as $\gamma > 1$ applies the converter is in the so-called conversion range. The range where $\gamma = 1$ is designated the clutch range since no further torque conversion takes place here.

Figure 8: Torque-conversion factor in hydrodynamic torque converter
1 Stall point,
2 Torque M_T at turbine or torque-conversion factor γ,
3 Drive torque M_P at impeller.
a Conversion range,
b Clutch range.
n_T Turbine speed,
n_P Impeller speed.



When the torque-conversion factor γ has dropped to 1, the oil flow acts on the stator vanes against the previous torque direction. The one-way clutch, which constitutes the connection between the stator and the transmission housing, transmits no torque in this direction. The stator now also rotates without torque in the oil flow. The co-rotation of the stator prevents an unwanted deflection of the oil flow in the clutch range and consequently a deterioration of the converter efficiency.

Lockup clutch
Even in the clutch range the turbine never quite reaches the speed of the impeller; a speed difference of 2 to 3 % remains. Without this slip no torque could be transmitted. To keep the losses in the converter at $n_T \approx n_P$ as low as possible, hydrodynamic transmission is locked up by a multi-disk clutch. Here the turbine shaft – i.e., the transmission input shaft – is coupled for example with the converter housing. To maintain the high vibration comfort of hydrodynamic isolation, one or more torsion dampers are incorporated in the lockup clutch's power-transmission path.

Advantages of the hydrodynamic torque converter
Because of the converter increase when driving off, the hydrodynamic torque converter is highly suitable for vehicles that have to tow heavy trailers (e.g., caravans). Further advantages are derived from this power take-up element not being subject to wear and the excellent ride comfort attributable to hydrodynamic torque transmission.

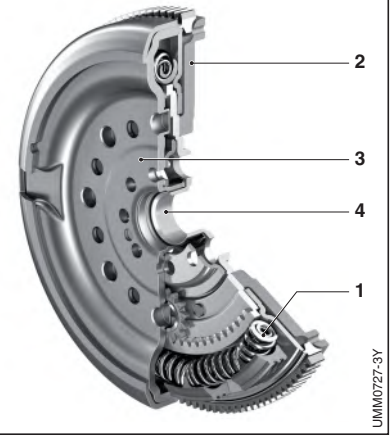
Vibration isolation

A significant requirement imposed on the drivetrain is comfort. The requirements imposed on vibration isolation are increasing continuously since optimizations in the combustion process of engines, smaller engine displacements with turbocharging, lower engine speeds in all driving ranges, and reduced numbers of cylinders do indeed lower fuel consumption but, due to high crankshaft radial accelerations, cause a higher rotational irregularity of the combustion-engine torque.

The purpose of vibration isolation is to reduce the torsional vibrations that occur as a result of the design of the internal-combustion engine. Vibration isolation can be effected by slip at the power take-up element, by damping the irregularity, by temporarily storing energy in a spring, by absorption, or by a combination of several methods.

In the case of vibration isolation by slip at the power take-up element, the mechanical energy of the vibration by friction is

Figure 9: Dual-mass flywheel with torsion damper (structural design)
1 Torsion damper (helical spring),
2 Primary flywheel with starter ring gear (from driving engine),
3 Secondary flywheel,
4 To power take-up element.



converted into heat. This energy is thus lost to the actual drive. Therefore this vibration-isolation option is now only used sporadically in critical vibration situations (e.g., low rotational speeds).

Torsion damper and dual-mass flywheel

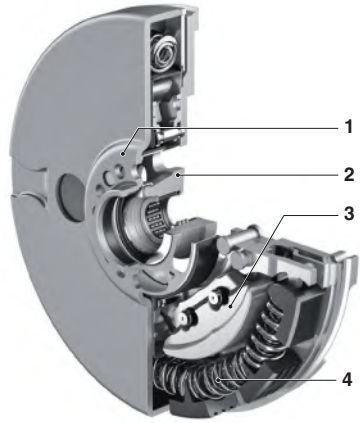
The torsion damper consists of coil springs arranged in a circle between the engine and the power take-up element (Figure 9), which during the vibration temporarily store energy in the compression of the springs (by a limited torsion between crankshaft and transmission inlet) and output this energy again to the transmission inlet. It is possible by varying the spring stiffness and diameter to adapt the torsion damper to the respective combination of engine and drivetrain.

In the case of the dual-mass flywheel, the mass of the engine flywheel is split and the torsion damper is placed between the two flywheel masses (Figure 9), which are pivoted against each other via a separate thrust and radial bearing arrangement. The moment of inertia on the transmission side is increased by apportioning to the primary flywheel mass on the engine side and to the secondary flywheel mass on the transmission side. Low spring stiffness can bring about a specifically "soft" harmonization of the spring units which connect the two flywheel masses to each other. As a result, the resonant frequency of the dual-mass flywheel is shifted significantly below the idle speed of the engine and the exciting engine orders. A natural oscillation of the dual-mass flywheel is avoided in this way.

Dual-mass flywheels are used in vehicles from the compact class upwards.

Figure 10: Centrifugal pendulum-type absorber and torsion damper

1 From driving engine,
2 To power take-up element,
3 Centrifugal pendulum-type absorber mass,
4 Helical spring (torsion damper).



Centrifugal pendulum-type absorber

In the case of the centrifugal pendulum-type absorber the rotational irregularity is balanced by means of a counterforce by vibrating masses additionally mounted in the torsion damper (Figure 10). Thanks to suitable lever conditions the vibrating masses can be selected at a proportionally small level compared with the mass of the internal-combustion engine to be isolated. The centrifugal pendulum-type absorber is speed-adaptive since the centrifugal acceleration on the vibrating mass increases as the rotational speed increases.

Centrifugal pendulum-type absorbers are more expensive to manufacture, but have the advantage of lower weight compared with dual-mass flywheels. They are used in vehicles from the compact class upwards.

References

[1] H. Naunheimer; B. Bertsche; J. Ryborz; W. Novak; P. Fietkan: Fahrzeuggetriebe – Grundlagen, Auswahl, Auslegung und Konstruktion. 3rd Edition, Springer-Verlag, 2019.
[2] P. Gerigk; D. Bruhn; J. Göbert: Kraftfahrzeugmechatronik – Nutzfahrzeugtechnik, 1st Edition, Westermann Schulbuchverlag, 2016.

Step-up gear units

Different gear ratios are implemented in the manual transmission for the purpose of adapting the torque made available by the engine and the power to the respective driving situation and to the driver's requests. Each gear ratio is called a gear or speed. Together with the transmission ratio of the final-drive differential the high engine speed is converted to the wheel speeds required for driving.

Manually shifted transmissions are a combination of step-up gear units and gear-shift elements. The vibration-isolation and the power take-up elements are arranged between the engine and the transmission (see Drivetrain).

For further references on transmissions, see [1] and [2].

Components in manually shifted transmissions

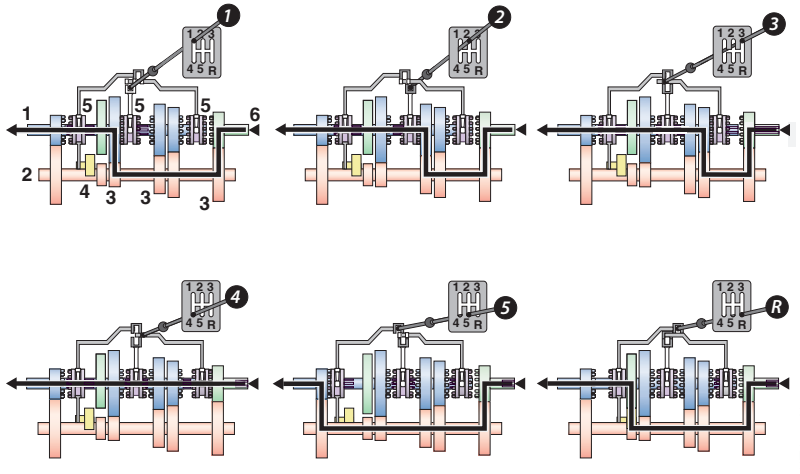
Step-up gear units

Gears

The gear (or gear wheel) is a machine element which has uniformly distributed teeth around its circular circumference. Torque and rotary motion are transmitted by a positive engagement occurring in succession of the teeth of a gear with the teeth of other gears. Different numbers of teeth on the gears – and thus different gear diameters (Figure 1) – determine the gear ratio from one gear to another. Rotational speeds and torques are inversely proportionally converted with this gear ratio, with the power output apart from efficiency remaining the same. The advantages of a step-up gear train lie in its high power density through slip-free transmission.

The type of toothing most commonly used in step-up gear units is involute

Figure 1: Design of a 5-speed manually shifted transmission with spur-gear sets
Depiction of the power flow in the various gears.
1 Main shaft (transmission output shaft to final-drive differential),
2 Countershaft, 3 Spur-gear sets for the different gears,
4 Idler gear for reverse gear, 5 Synchromeshes,
6 Transmission input shaft.
1, *2*, *3*, *4*, *5* Forward gears,
R Reverse gear.



toothing. Here the cross-section of a tooth flank depicts a part of an involute of a circle (Figure 2). This ensures that the engaging teeth touch each other along a straight line of action. The aim of involute toothing is to achieve a high amount of rolling and an accordingly low amount of sliding during the transmission of torque in order to minimize wear and heat generation in the toothing. Further advantages of involute toothing are the high tolerance in relation to wheelbase tolerances and relatively simple manufacturability. Figure 3 shows the parameters for a gear.

Figure 2: Involute of a circle
1 Base circle,
2 Involute,
d_b Base-circle diameter.

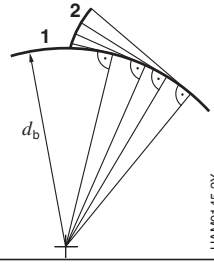
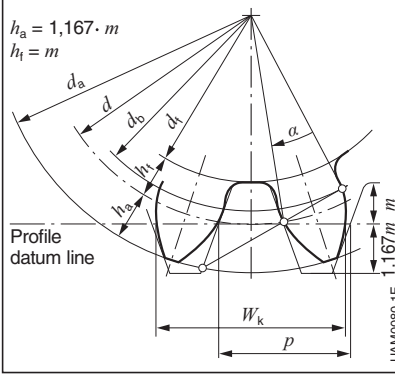


Figure 3: Parameters for a gear
d_a Outside diameter,
d Pitch diameter,
d_b Base-circle diameter,
d_f Root diameter,
h_a Addendum,
h_f Dedendum,
W_k Base tangent length,
p Pitch $p = \pi m$,
m Modulus,
α Pressure angle.

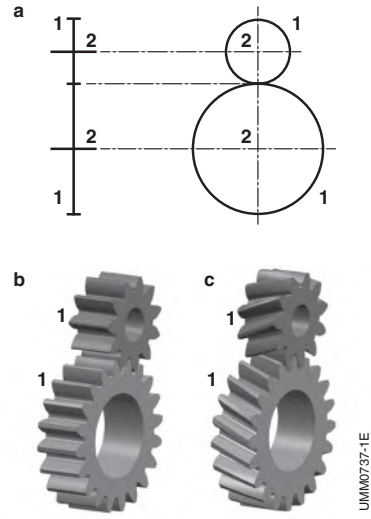


Straight and helical toothing

Helical gears are used for the most part in spur-gear drives. On these gears the teeth are arranged not parallel with the gear axes, but at an angle to them (Figure 4c). When the teeth of two gears mesh, the teeth do not bear on their full face width from the outset, as is the case with straight spur gears (Figure 4b). Instead, the bearing face width increases slowly as the gears continue to rotate until the teeth have contact over their full width. As the gears rotate out of the action zone the face width subjected to load also decreases slowly. In this way, the kinetic impulses when meshing are much smaller than on straight gears. Accordingly, fewer oscillations are excited, which provides for quieter running of the gears.

Acoustically, it is possible to discern the difference between straight and helical toothing in older manually shifted transmissions. In such transmissions the reverse gear is straight-toothed and "whines" when reversing, while the forward gears are acoustically inconspicuous when driving

Figure 4: Spur-gear set
a) Schematic diagram,
b) Structural design of a straight spur-gear set,
c) Structural design of a helical spur-gear set.
1 Spur gear, 2 Axis.



forwards. Another advantage of helical gears is their higher load capability in that more teeth are simultaneously engaged than on straight gears.

Helical toothing generates axial forces which must be absorbed by means of a suitable bearing arrangement. Moreover, they are more complex to produce than straight spur gears.

Spur-gear set

In spur-gear sets the gears are arranged on parallel axes (Figure 4). One of the gears is driven while the other is coupled to the output. The speed ratio from the drive gear to the output gear is obtained from the different numbers of teeth. In the vehicle transmission the gears can either be permanently coupled on shafts (fixed gear) or pivoted on them (movable gear).

Spur gears are easier to manufacture than internally toothed ring gears of a planetary-gear set. Moreover, the axial offset between input and output suggests itself for front-transverse drivetrains.

Spur-gear sets are used in manually shifted transmissions, automated manual transmissions, and dual-clutch transmissions.

Planetary-gear set

The planetary-gear set consists of a sun gear, an internally toothed ring gear, and several planetary gears. The planetary gears are interconnected by a carrier and mounted on this (Figure 5). The planetary-gear set, in contrast to the spur-gear set, has not only fixed axes. The axes of the planetary gears can circulate in an orbit around the axes of the sun and ring gears.

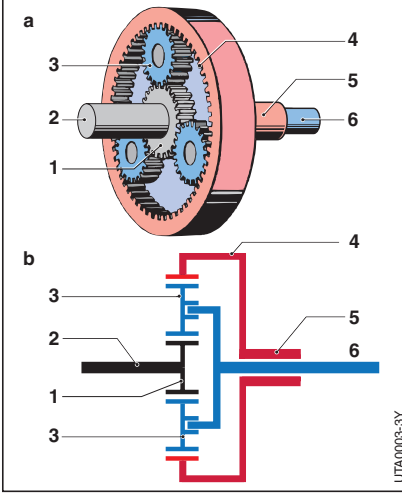
Basically, in each case one of the three elements (sun gear, ring gear, carrier) in the planetary-gear set can be used as the drive, while a second element is permanently joined to the transmission housing in order to provide output via a third element. With three drive options and three output options (with the third element stationary in each case) there are six different gear ratios which a planetary-gear set can offer when connected appropriately. Furthermore, by joining the three elements to each other, it is possible to achieve a circulation of the planetary-gear set as a block with the gear ratio one as the seventh ratio. However, the ratios of a planetary-gear set are not so staged that a planetary-gear set alone would be usable for the desired ratios of a passenger-car transmission.

Thanks to the symmetrical design and the distribution of torque to several planetary gears, the power density in the planetary-gear set is particularly high and accordingly the size in relation to the transmitted torque is small. Because of the coaxial position of the input and output shafts this type is particularly suitable for standard drivetrains. Several planetary-gear sets combined via shift elements are predominantly used in automatic planetary transmissions.

A planetary-gear set can also be used as a summation gearbox or transfer case. In this case, none of the planetary-gear set elements are joined to the housing. When it is used as a summation gearbox, two elements of the planetary-gear set are driven and output via the third element. When it is used as a transfer case, one of the planetary-gear set elements operates as the drive while the other two operate as the output.

Figure 5: Planetary-gear set

a) Structural design,
b) Diagram.
1 Sun gear, 2 Shaft for sun gear,
3 Planetary gears, 4 Ring gear,
5 Hollow shaft for ring gear,
6 Shaft with planetary-gear carrier.



Gear-shift elements

Function of synchromesh

Manually shifted transmissions, automated manual transmissions, and dual-clutch transmissions as a rule have for each gear a spur-gear set whose teeth are permanently engaged (Figure 1). In each case one gear of the spur-gear set is permanently mounted on the shaft while the other is pivoted on it. In Figure 1 the gears with the exception of the first gear are pivoted on the main shaft (movable gears, control gears). The opposing gears in each case on the countershaft are permanently joined to the shaft and are thus not pivoted on it (fixed gears). The control gears are all permanently driven by the opposing fixed gear in each case. When a gear is not engaged, the control gears consequently rotate at a speed different from the shaft.

In the event of a gear change the corresponding control gear is sped up or slowed down to the shaft speed. When synchronism is in place a non-positive connection with the shaft is established. The device that brings about the equal rotational speeds is called a synchromesh.

Design

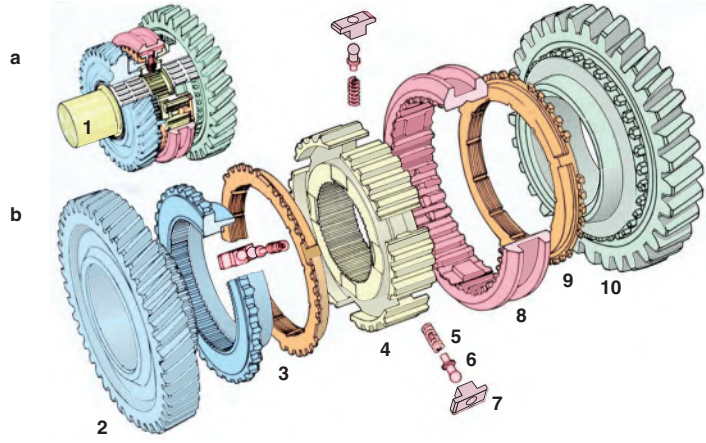
The synchronizer body is positively connected with the shaft (Figure 6). It carries the axially movable gearshift sleeve. The synchronizer body has recesses for three sliding blocks, which are pressed outwards by two compression springs against the gearshift sleeve. The gearshift sleeve features on the inside dog toothing which has recesses for the sliding blocks. In this way, it is possible to slide the gearshift sleeve with the sliding blocks to the right and left.

Synchronizing action

In the course of the synchronizing action the process of sliding the gearshift sleeve with the sliding blocks moves the synchronizer ring towards the conical friction surface on the control gear to be shifted (Figure 6). Recesses are worked into the synchronizer ring for the sliding blocks; the sliding blocks carry the synchronizer ring. The process of pressing the synchronizer ring with force F_S onto the conical friction surface of the control gear generates by way of sliding friction a torque M_S which would like to match the control gear to the speed of the shaft. Lockout toothing on the

Figure 6: Synchromesh

a) Assembly,
b) Exploded view.
1 Main shaft, 2 Movable control gear, 3 Control ring with clutch teeth,
4 Synchronizer body (permanently mounted on main shaft), 5 Compression spring,
6 Catch pin, 7 Sliding block, 8 Gearshift sleeve with dog toothing,
9 Synchronizer ring with inner cone and lockout teeth,
10 Movable control gear with synchronizer cone and shift toothing.

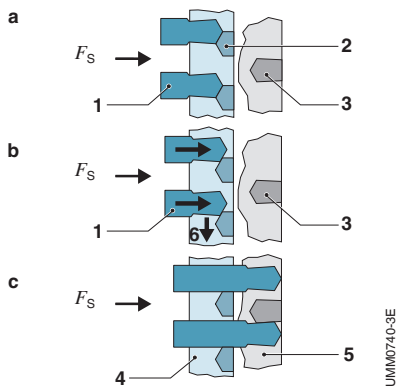


UMM0739-2Y

synchronizer ring prevents the dog toothing of the gearshift sleeve from meshing with the shift toothing before the control gear and shaft are synchronized (Figure 7). In this process the dog toothing of the gearshift sleeve contacts the tooth angles of the lockout toothing on the synchronizer ring and the friction torque of the synchronizer ring counteracts any evasion of the lockout toothing. Further rotation of the synchronizer ring relative to the gearshift sleeve is prevented by axially movable sliding blocks.

The recesses for the sliding blocks in the synchronizer body are exactly the same width as the sliding blocks, while the recesses in the synchronizer ring are slightly wider. In this way, the synchronizer ring can, as pictured in Figure 7, move relatively through a small angle. The sliding blocks transmit the torque from the synchronizer body to the synchronizer ring. The external teeth on the synchronizer body, synchronizer ring and shift toothing are dimensioned in such a way that they can establish positive locking with the dog toothing of the gearshift sleeve.

Figure 7: Synchronizing action
Force F_S on the synchronizer ring.
a) Synchronization and lockout,
b) Rotate synchronizer ring back,
c) Establish positive locking.
1 Dog toothing on gearshift sleeve,
2 Lockout toothing on synchronizer ring,
3 Shift toothing on control gear,
4 Synchronizer ring,
5 Control gear,
6 Relative motion of synchronizer ring in relation to gearshift sleeve.



Once the synchronization of the control gear and the shaft is achieved, the synchronizing torque and consequently also the torque acting on the lockout toothing are zero. The applied shift force F_S now induces, via the tooth angles of the synchronizer-ring lockout toothing, the synchronizer ring to rotate relative to the gearshift sleeve so that the shift-sleeve dog toothing can be slid over the synchronizer-ring lockout toothing into the control-gear shift toothing. This establishes positive locking between the shaft and the control gear.

Increased shift force can reduce the synchronizing time, but in the long term this will cause increased wear of the friction lining on the synchronizer ring. Also, the lockout toothing can be turned back by increased shift force already before synchronization and the shift-sleeve dog toothing advanced up to the control-gear shift toothing. When the shift-sleeve dog teeth and the shift toothing on the control gear engage at different speeds, this results in “gear-shift rasping”.

Multi-cone synchromesh

Synchromeshes with two or three conical friction surfaces arranged on top of each other are used to synchronize components with high mass inertia. These are referred to as double-cone or triple-cone synchromesh.

Often two and sometimes even three characteristic forms of synchromesh are fitted in transmissions. For the most part the control gears of first gear and reverse gear are equipped with triple-cone synchromesh because here the differences in speed between the gearshift sleeve and the loose running control gear are greater than in higher gears. The control gears of second and third gears are equipped with double-cone synchromesh, and the other gears with single-cone synchromesh. The advantage of this arrangement in manually shifted transmissions is that the shift forces are similarly high for all the gears.

Multi-disk clutch

A multi-disk clutch usually consists of a number of inner and outer disks. The inner disks have on their inside diameter synchronization toothing for example for joining the clutch to components in the transmission, while the outer disks have the synchronization toothing on their outside diameter (Figure 8). A disk is glued in each friction contact with a paper-like lining; as a rule this is the disk with the internal toothing. The other disk has a smooth steel surface.

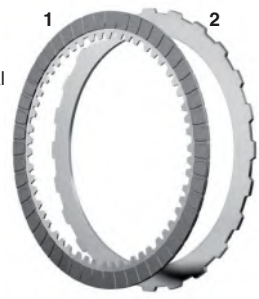
The disks are axially movable thanks to the connection via the toothing. They can be pressed together by means of a hydraulic mechanism (Figure 8) so as to create non-positive connection.

A higher torque can be transmitted for the same operating force by the use of a number of alternately stacked inner and outer disks and the greater lining surface area thereby obtained. To ensure that the multi-disk clutch is efficiently cooled, the paper lining has grooves for a constant throughflow of cooling oil.

The oil-cooled starting clutch and the dual clutch are distinct types of the multi-disk clutch. Multi-disk clutches are of fundamental importance to the function of automatic planetary transmissions, in which the gear to be shifted is introduced into the energy flow with frictional engagement and without interruption of tractive force.

Figure 8: Multi-disk clutch

1 Inner disk with internal toothing,
2 Outer disk with external toothing.

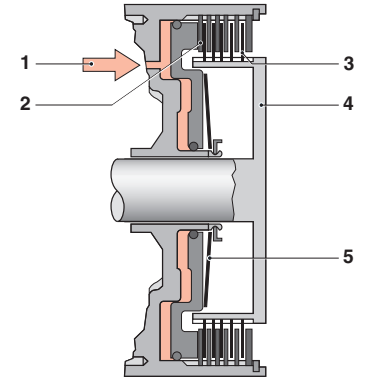


Multi-disk brake

The multi-disk brake is a special type of multi-disk clutch. The multi-disk brake differs in that for example the outer disks are connected to a non-rotating component such as the transmission housing and the clutch disks are stationary in the closed state while the elements of a multi-disk clutch rotate in the closed state. The multi-disk brake can be used for example to support a planetary-gear set element against the transmission housing in order to effect a transmission ratio.

Figure 9: Multi-disk clutch

1 Compressed-oil feed,
2 Outer disk,
3 Inner disk,
4 Disk carrier,
5 Return spring.



Manually shifted transmissions

Design

Manually shifted transmissions (MTs) are usually designed as spur-gear drives in which the gear shift and the drive-off process are performed manually by the driver. Normally a dry single clutch is used as the power take-up element for passenger-car applications, double-disk clutches are also used for heavy commercial vehicles. In manually shifted transmissions each gear – except for transmissions with direct gear, in which the input shaft is directly connected to the output shaft – is rendered by a spur-gear pair (Figure 10). The different ratios of the numbers of teeth per spur-gear pair produce the different gear ratios.

In a double-shaft transmission the gears are seated on the main shaft or the countershaft. All the gears are permanently mounted on these shafts and as a result all the spur-gear pairs are engaged. Each spur-gear pair of a gear has a fixed gear, which is permanently connected to a shaft, and a movable gear (control gear), which is pivoted on the other shaft. The control gears can be seated on the main shaft, but also on the countershaft. For each gear the associated control gear is non-rotationally connected to the shaft by gear-shift elements. The individual control gears are coupled via synchromeshes (see Figure 6). The gearshift sleeves are moved into the position appropriate for the gear to be selected via gearshift forks and gearshift rods depending on the gearshift lever (Figure 11).

Figure 11: Control gear with gearshift sleeve and gearshift equipment

1 Main shaft, 2 Gearshift rod, 3 Gearshift fork,
4 Control gear, 5 Control ring,
6 Gearshift-sleeve toothing,
7 Gearshift sleeve.

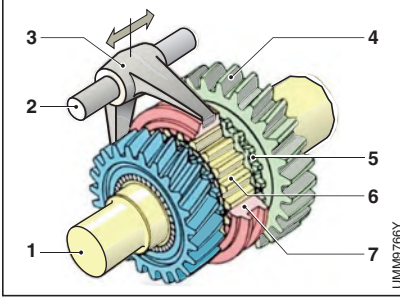
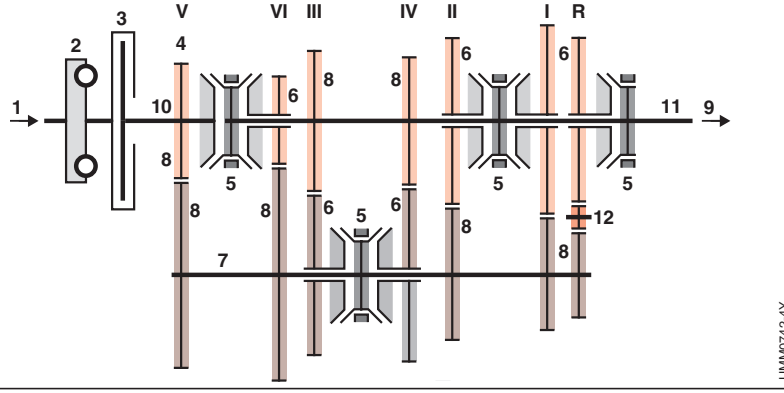


Figure 10: Schematic diagram of a 6-speed manually shifted transmission for a standard drive

1 From driving engine, 2 Dual-mass flywheel, 3 Clutch,
4 Constant transmission ratio (5th gear), 5 Synchromeshes, 6 Control gear,
7 Countershaft, 8 Fixed gear, 9 Transmission output to final-drive unit,
10 Transmission input shaft (input shaft), 11 Transmission output shaft,
12 Idler gear for reverse gear.
I…VI First through sixth gear, R Reverse gear.



Manually shifted transmissions for the most part have five to six forward gears and one reverse gear. The spacing (ratio of the gear ratios from the lowest to the highest gear) of the transmissions is roughly four to eight.

Manually shifted transmissions belong to the category of tractive-force-interrupted transmissions. To change a transmission step the power flow must be disconnected by the clutch.

Transmissions for standard drive

The designs for a standard drive (Figure 10 and Figure 12) have a coaxial input and output and a countershaft. In other words, the input shaft (transmission input shaft) and output shaft (main shaft) lie in the same plane (transmission with in-plane shafts). The direction of the power flow is not altered. The direction of rotation of the output shaft is the same – apart from for the reverse gear – as the direction of rotation of the input shaft because the power flow is provided via spur-gear pairs or in direct gear without transmission ratio by gears.

A constant transmission ratio transmits the engine power from the transmission input shaft to the countershaft. The control gears of the different transmission ratios are coupled by means of synchromeshes to the main shaft or countershaft so that the power flow can be delivered via the spur-gear pair in question to the transmission output.

The direction of rotation is reversed for the reverse gear by an idler gear.

Transmissions for front-wheel drive

Types with one countershaft (double-shaft transmission, see Figure 13) or two countershafts parallel to the drive (triple-shaft transmission) have established themselves in front-wheel drives. In a front-wheel drive the final-drive unit with differential is integrated in the transmission.

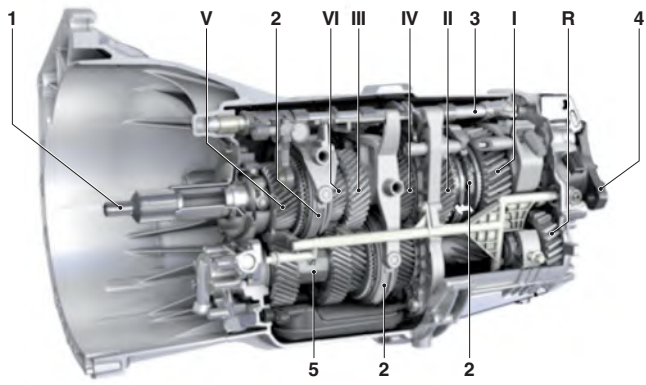
Double-shaft transmissions

In the double-shaft transmission in Figure 13 the fixed gears are seated on the input shaft and the control gears are non-rotationally connected to the countershaft according to the selected gear. This is a transmission without out-of-plane shafts because the input and output shafts do not lie in the same plane. This is because only one spur-gear pair is used to establish a transmission ratio.

The final-drive unit is driven in the double-shaft transmission by a constant transmission ratio of the countershaft.

Figure 12: Structural design of a 6-speed manually shifted transmission for a standard drive

1 Transmission input shaft, 2 Synchromeshes, 3 Gearshift linkage,
4 To final-drive unit, 5 Countershaft.
I…VI First through sixth gear,
R Reverse gear.



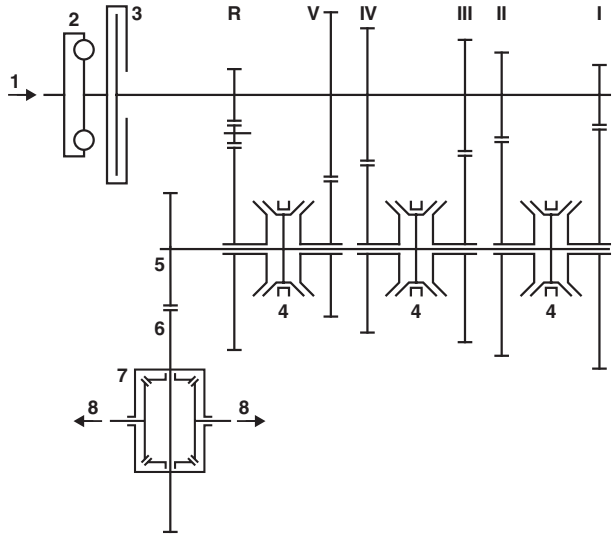
Triple-shaft transmissions
Splitting the transmission steps between two parallel countershafts enables the transmission to be short in its construction. The gears of the transmission input shaft drive the gears on the two countershafts (transmission output shafts). Seated on these shafts are the synchromeshes, which couple the gear responsible for the selected gear to the corresponding countershaft. In each case one gear on the countershafts meshes with the spur gear of the differential.

Depending on the gear selected, one gear or the other is non-rotationally coupled with the corresponding countershaft and drives the spur gear of the differential. The space required for components is reduced thanks to the control gears and the synchromeshes being split between two countershafts.

The overall length can be further optimized for a triple-shaft transmission by the double use of fixed gears on the drive shaft. Here one fixed gear meshes with two control gears (e.g., for the third and fifth gears) on the different countershafts.

Application
Manually shifted transmissions are still popular in Europe even in new vehicles. In the USA and Asia the demand for manually shifted transmissions is declining, apart from in certain sports cars.

Figure 13: Schematic diagram of a 5-speed transmission for front-transverse drive
1 From driving engine, 2 Torsion damper, 3 Clutch,
4 Synchromeshes, 5 Output constant, 6 Differential spur gear,
7 Differential, 8 To wheels.
I…V First through fifth gear,
R Reverse gear.



Automated manual transmissions

Design

Automated manual transmissions (AMTs) have a mechanical design comparable to a manually shifted transmission (Figure 13).

Operation

The clutch is operated by electromotive or electrohydraulic actuators (Figure 14) which are actuated by a transmission control unit. Figure 15 shows the basic design of an electromechanical system. As is the case for operating the clutch, servomotors are used to operate the synchromeshes and thereby change gear. This helps to provide the operating comfort of a fully automatic transmission. The driver can choose between fully automatic control and manual input of the gear. There is, however, no clutch pedal.

With the electrohydraulic control system, actuation is essentially performed via an electrically driven hydraulic pump, a pressure accumulator, and a hydraulic actuation unit with the actuators for selecting the gears.

Figure 14: Structural design of an actuator for an automated manual transmission for front-transverse drive

1 Clutch,
2 Clutch actuator.



Advantages and disadvantages

Like the manually shifted transmission, the automated manual transmission is a tractive-force-interrupted transmission, which gives rise to disadvantages in shifting comfort compared with other automatic-transmission construction types. However, the interruption of tractive force is generally shorter than in a manually shifted transmission.

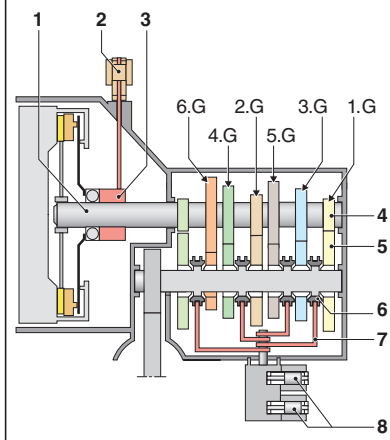
Efficiency is higher when compared with an automatic planetary transmission, resulting in lower fuel consumption.

Application

Automated manual transmissions are used as standard in the commercial-vehicle sector, but have not been able to make the breakthrough in the passenger-car sector. The reason for this is the poor comfort due to the interruption of tractive force during the gearshift.

Figure 15: Basic design of an automated 6-speed transmission for front-transverse drive

1 Input shaft,
2 Electronic clutch management,
3 Central disengaging gear,
4 Fixed gears,
5 Pivoted control gears,
6 Gearshift sleeve,
7 Gearshift linkage,
8 Servomotors.



Dual-clutch transmissions

Design

Dual-clutch transmissions (DCTs), also known as direct-shift transmissions (DSTs), consist of two separate spur-gear sub-transmissions which can be connected in each case via a clutch (see Double clutch) to the internal-combustion engine.

The two clutches in the double clutch can be operated independently of each other. They can be designed as dry-running or wet-running systems (see Double clutch). The wet-running clutch is used if high torques are to be transmitted. A hollow shaft leads from the inner clutch to the transmission; the solid shaft driven by the outer clutch runs inside the hollow shaft. The gear wheels for the individual gears are seated as fixed gears or as control (movable) gears on these two shafts (Figure 16, Figure 17).

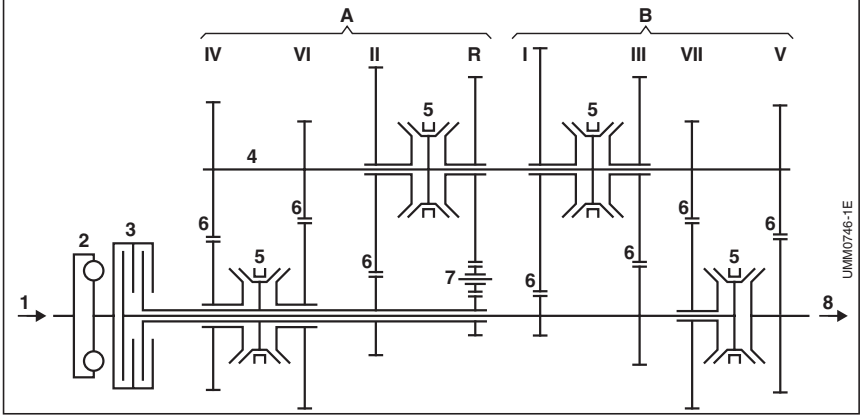
The gears are usually lubricated by direct splash oil. The oil is supplied by a mechanically driven oil pump or an electric oil pump. The oil cooler optionally mounted on the transmission is integrated in the cooling circuit.

Operation

By splitting the even gears (two, four, six) and odd gears (one, three, five, seven) between the two sub-transmissions (Figure 16) it is possible to change the gear ratios by means of power-shifting. When sub-transmission A is engaged via the inner clutch, in sub-transmission B the next gear higher or the next gear lower is pre-assigned via the corresponding synchromesh. The electronic transmission control uses among others the accelerator-pedal position, the engine load, and the engine speed to decide which gear is pre-assigned. For example, a fully depressed accelerator pedal and a low engine speed indicates that the next gear lower must be selected (kickdown). When the inner clutch is released and at the same time the outer clutch closes (or vice versa), the gearshift to the pre-assigned gear is initiated. The clutches are actuated by clutch valves. The gearshift sleeves are controlled by hydraulically actuated gearshift forks.

The overlap of closing and opening of the clutches results in a nearly tractive-force-interruption-free gear change. A dual-clutch transmission therefore offers the comfort

Figure 16: Schematic diagram of a 7-speed dual-clutch transmission for a standard drive
1 From driving engine, 2 Dual-mass flywheel, 3 Dual clutch, 4 Countershaft,
5 Synchromeshes, 6 Spur-gear sets, 7 Idler gear for reverse gear,
8 To final-drive unit.
I…VI First through seventh gear, R Reverse gear.
A Sub-transmission with even gears,
B Sub-transmission with odd gears.



of an automatic transmission. But only one gear on the other sub-transmission can ever be preselected. However, multiple upshifts or multiple downshifts can be achieved sequentially or by means of so-called support shifts via an intermediate gear (power transmission by slipping clutch during the gearshift).

The reverse gear can be assigned to either the even or the odd sub-transmission. In Figure 16 the reverse gear is assigned to the even sub-transmission. The direction of rotation is reversed by an idler gear.

All the synchromeshes are moved into the center position once Neutral has been selected with the gearshift lever. As a result no gear is engaged. When the engine is stopped the two clutches are open. The transmission therefore needs a parking lock. A parking-lock gear engages the spur gear of the final-drive unit. A locking pawl blocks the parking-lock gear when the gearshift lever is moved into the P position. In the Park position the mechanical parking lock stops the vehicle from rolling away.

The clutches are used both for load sharing during shifting and for the drive-off process.

Advantages and disadvantages

The advantage of dual-clutch transmissions is that, unlike automatic planetary transmissions, they are suitable for engine speeds in excess of 7,000 rpm. Dual-clutch transmissions are therefore used in among others sports cars.

The spur-gear construction in dual-clutch transmissions facilitates advantageous gear-set versions especially in front-transverse transmissions with their axial offset between transmission input and transmission output shafts. In these gear sets the final-drive differential, unlike an automatic planetary transmission, is driven directly by the countershafts, providing for less meshing and consequently greater transmission efficiency in front-transverse transmissions.

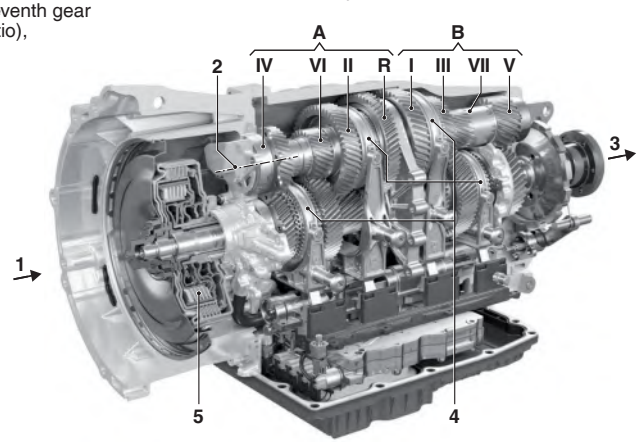
In contrast to a transmission with a hydrodynamic torque converter there is no torque multiplication in a dual-clutch transmission when driving off. Therefore a greater transmission ratio and thus a greater transmission spacing are required in first gear.

Application

Thanks to the dual-clutch transmission the automatic transmission has also found its way into the lower vehicle classes with smaller engines and front-wheel drive.

Figure 17: Structural design of a 7-speed dual-clutch transmission for a standard drive

1 From driving engine, 2 Countershaft, 3 To final-drive unit, 4 Synchromeshes, 5 Dual clutch.
I…VII First through seventh gear (fifth gear as direct ratio),
R Reverse gear,
A Sub-transmission with even gears,
B Sub-transmission with odd gears.



Automatic planetary transmissions

Design

Automatic planetary transmissions (ATs) realize their transmission steps through a series of planetary-gear sets arranged in succession which are connected to each other in different ways by shift elements (multi-disk clutches and mullti-disk brakes (Figure 18). The number and arrangement of planetary-gear sets and clutches are dependent on the number of gears implemented in the transmission.

A hydrodynamic torque converter usually serves as the power take-up element for an automatic planetary transmission.

Operation

Depending on the shift system, different transmission ratios can be set with a single planetary-gear set; however, not all these ratios can be used for driving (see Planetary-gear set). Therefore a number of planetary-gear sets are required. An 8-speed transmission can for example be achieved with four planetary-gear sets and five shift elements (Figure 18). By means of different combinations of several simultaneously closed or opened shift elements the transmission schematic shown (Table 1) can realize eight forward gears and one reverse gear.

Planetary-gear sets are often used in which parts of the individual single planetary-gear sets are structurally brought together. Two types of planetary transmissions have gained acceptance. In the Ravigneaux gear set two different planetary-gear sets and sun gears operate in a common ring gear (Figure 19); in the Simpson gear set two planetary-gear sets and ring gears operate around a common sun gear. These designs reduce structural complexity and costs.

In automatic planetary transmissions eight to nine gears have proven to be the optimum with regard to reduced fuel consumption and good driveability. For a smaller number of gears the combustion-engine program map is not yet opti-

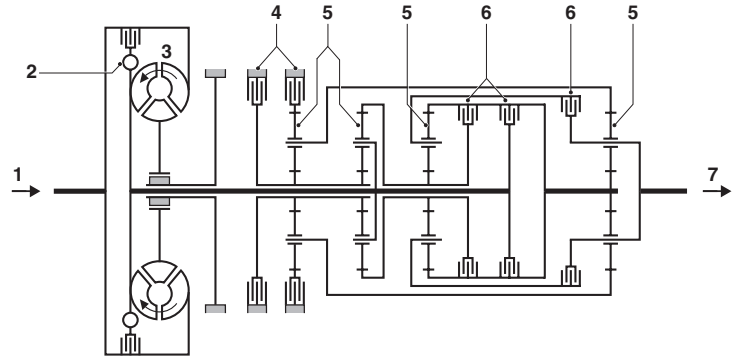
Table 1: Layout of gears of an 8-speed automatic planetary transmission

Gear	Brake		Clutch			Ratio
	A	B	C	D	E	
1	+	+	+	0	0	4.714
2	+	+	0	0	+	3.140
3	0	+	+	0	+	2.106
4	0	+	0	+	+	1.667
5	0	+	+	+	0	1.285
6	0	0	+	+	+	1.000
7	+	0	+	+	0	0.839
8	+	0	0	+	+	0.667
R	+	+	0	+	0	–3.317

\+ Shift element activated
0 Shift element deactivated

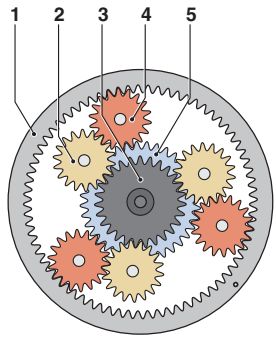
Figure 18: Schematic diagram of an 8-speed automatic planetary transmission for a standard drive

1 From driving engine, 2 Torsion damper, 3 Hydrodynamic converter, 4 Multi-disk brakes, 5 Planetary-gear sets, 6 Multi-disk clutches, 7 To final-drive unit.



mally developed to drive constantly in the internal-combustion engine's optimum consumption and performance range. For a larger number of gears additional shift elements or planetary-gear sets are required which diminish the efficiency of the transmission and hence the drivetrain in most driving situations. The energy that is required for the greater number of gear shifts can no longer be compensated by better utilization of the combustion-engine program map.

Figure 19: Ravigneaux gear set (diagram).
1 Ring gear, 2 Planetary-gear set 1,
3 Sun gear 1, 4 Planetary-gear set 2,
5 Sun gear 2.



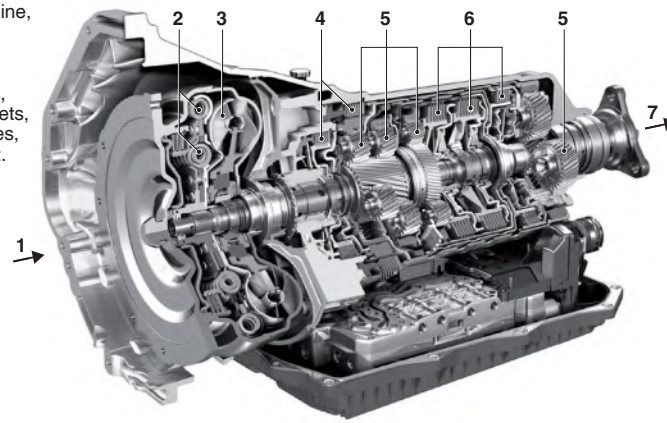
Advantages and disadvantages
The planetary-gear sets have coaxial input and output. This results in advantageous transmission construction types particularly for the standard drive, where the engine crankshaft and the axle-differential drive are on the same axis (Figure 20).

Since the transmitted power is shared in the planetary-gear sets between a number of planetary gears spread over the circumference, automatic planetary transmissions are characterized by a high power density. The comfort of a power-shift transmission is achieved by targeted overlapping shifts of the clutches.

The hydrodynamic torque converter provides for high driveaway and maneuvering comfort. The torque multiplication in the converter has a positive effect on the vehicle's driveaway performance. The slip in the converter is eliminated with a converter lockup clutch. This reduces the losses in this transmission and makes it efficient, even in comparison with manual transmissions.

Application
Automatic planetary transmissions, owing to their type of construction, are used in vehicles with longitudinally installed engines, i.e., in mid-range vehicles and upwards, but also in buses.

Figure 20: Structural design of an 8-speed automatic planetary transmission for a standard drive
1 From driving engine,
2 Torsion damper,
3 Hydrodynamic converter,
4 Multi-disk brakes,
5 Planetary-gear sets,
6 Multi-disk clutches,
7 To final-drive unit.



Continuously variable transmissions

Power and torque transmission via variator

In continuously variable transmissions (CVTs) power is transmitted with a variable gear ratio via a wrap belt (pushbelt) or a wrap chain which runs in a groove from a first to a second V-pulley set (Figure 21). The combination of two V-pulley sets and a wrap belt or a wrap chain is called a variator. Each V-pulley set consists of two tapered pulleys which facing each other. One of these pulleys is axially movable ("loose"). This axial displacement changes the spacing between the pulleys and with it the radius on which the pushbelt or the chain runs. The combination of different, infinitely variable running radii on the two V-pulley sets products the variator's gear ratio.

The transmission of torque from one V-pulley set to the other is based on friction. The required frictional force is achieved by hydraulically pressing together the pulleys of each V-pulley set. Each loose pulley contains one or two hydraulic pistons to which oil pressure is applied by the control system. The amount of pressure must be determined for each V-pulley set. The force that must be exerted on the pushbelt or the chain depends on the instantaneous torque, the gear ratio, and the desired gear ratio. The pressure is then calculated taking account of the required force, the surface of the piston (or pistons), and the actual rotational speed of the V-pulley set. The centrifugal-force effect also generates a pressure which must be taken into account.

Design of a CVT

Figure 21 shows the typical mechanical design of a CVT. On the input side a power take-up element is used, usually a torque converter (as pictured in Figure 21), or less commonly a wet multi-disk clutch. A planetary-gear set is typically used to shift between forward gear, neutral, and reverse gear. A final-drive unit (usually in the form of two transmission steps) and a differential are arranged after the variator. Figure 21 does not show the control system and the parking mechanism.

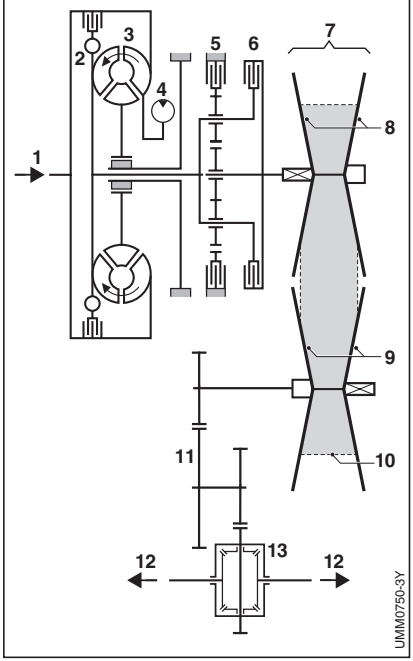
Because of the relatively large center distance between the variator input and output, CVTs are best suited for use in a front transverse drive (Figure 22). However, they are also used in combination with longitudinally installed engines.

Open-loop control

Actuation, cooling and lubrication of the CVT is performed by an electronically controlled hydraulic system. The electronic control system uses a number of sensors and actuators in the transmission and information from other systems in the vehicle which is normally transmitted via a CAN bus.

Figure 21: Schematic diagram of a continuously variable transmission

1 From driving engine, 2 Torsion damper,
3 Hydrodynamic torque converter with lockup clutch,
4 Oil pump, 5 Reverse-gear brake,
6 Forward-gear clutch, 7 Belt variator,
8 First V-pulley set (input),
9 Second V-pulley set (output),
10 Belt or chain, 11 Final-drive differential,
12 To the wheels, 13 Differential.



The hydraulic unit contains values that control the oil pressure or the flow to a number of components in the CVT. The oil is supplied by an engine-driven pump, usually a gear pump or a vane pump. In the last few years more and more CVTs are being equipped with an additional (small) electric oil pump which keeps the hydraulic system filled when the engine is shut down by a start/stop system. The oils used in CVTs are special oils similar to ATF which are optimized in terms of their friction properties.

CVT features

Because the gear changes take place with continuous variability, the CVT offers a high level of ride comfort without tractive-force interruption. Unusual driveability as experienced by some drivers can be avoided by using under certain circumstances programmed “fixed” gear ratios.

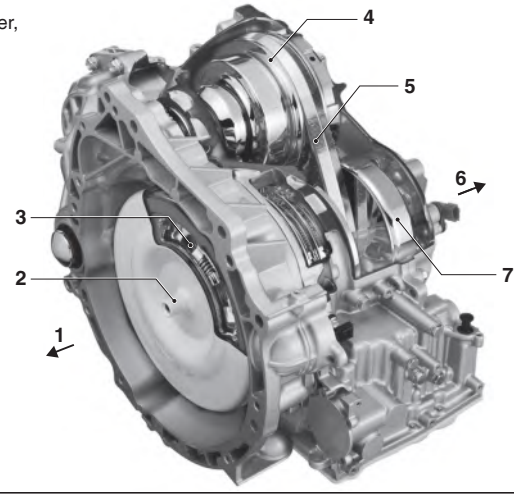
A transmission spacing with a factor of up to 7.3 enables the engine to be operated in a wide speed and load range, which is beneficial in terms of both power output and fuel consumption. This benefit outweighs the naturally higher losses of a friction-based transmission. With regard to mechanical efficiency, the wrap chain fares slightly better than the pushbelt. However, the pushbelt is better when it comes to NVH (Noise Vibration Harshness).

Applications

CVTs are used in passenger cars with both internal-combustion engines and hybrid drives. CVTs for electric vehicles are being developed. CVTs for vehicles with lower power output such as mopeds, snowmobiles and side-by-side vehicles (SSVs) frequently use a dry rubber-belt variator and a mechanical control system.

Figure 22: Structural design of a continuously variable transmission (source: Jatco)

1 To wheel,
2 From driving engine,
3 Hydrodynamic torque converter,
4 Second V-pulley set,
5 Wrap belt (pushbelt),
6 To wheel,
7 First V-pulley set.



Commercial-vehicle transmissions

Commercial-vehicle transmissions differ significantly in their dimensions from passenger-car transmissions. The reasons for this are not only higher engine torques and engine power output, but also above all the much higher mileage demands of sometimes over one million kilometers. The commercial-vehicle sector features predominantly manually shifted transmissions, automated manual transmissions, and automatic planetary transmissions.

Manually shifted transmissions and automated manual transmissions

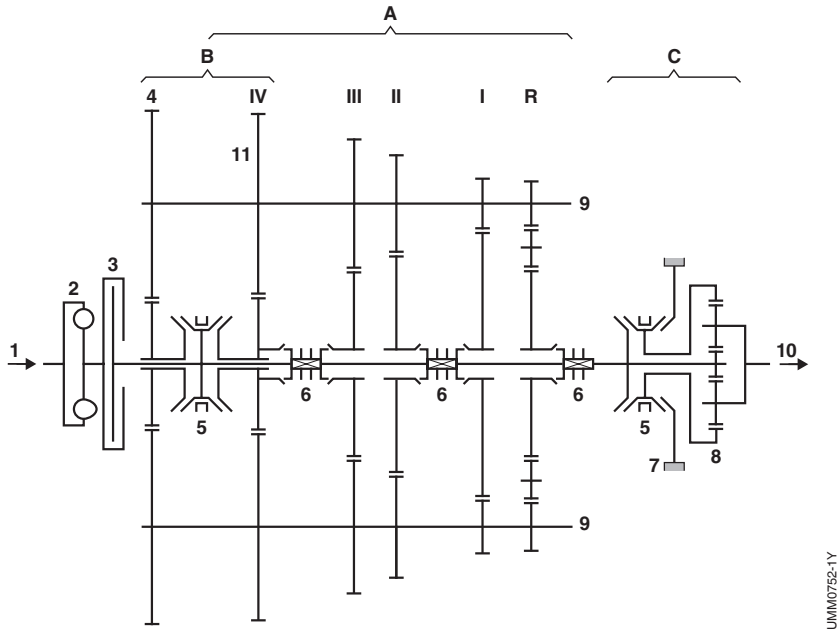
The standard construction has gained acceptance as the drivetrain concept in trucks with a gross weight rating of over 3.5 tons. The high loads on the rear powered axle(s) and the associated potential high drive force of the drive wheels are the reason for this drivetrain concept.

Sub-transmissions

Primarily manually shifted transmissions and automated manual transmissions have become prevalent in the truck sector. Up to 16 gears are obtained in manually shifted transmissions and automated manual transmissions. To minimize construction expense, two to three sub-transmissions (splitter box, main box, range-change box) are arranged in succession (Figure 23). In a truck transmission with 16 gears the main box has for example four forward gears and one reverse gear. Mounted upstream of this main box is a splitter box with two gears. The splitter box allows intermediate ratios

Figure 23: Schematic diagram of a 16-speed truck range-change transmission with splitter box and range-change box

1 From driving engine, 2 Torsion damper, 3 Clutch, 4 Constant transmission ratio 1,
5 Synchromesh, 6 Shift dogs, 7 Housing connection, 8 Planetary-gear set,
9 Countershafts, 10 To final-drive unit, 11 Constant transmission ratio 2 (fourth gear).
I…IV First through fourth gear in main box.
R Reverse gear.
A Main box, B Splitter box, C Range-change box.



of the four gears of the main box. Often the splitter box is used with a full payload since only small gear steps are possible for the very high vehicle weights in relation to the drive power. When the truck is unladen, the ratio of engine power to vehicle weight is much more favorable such that the greater gear steps of the main box can be shifted without using the splitter box.

Splitter and main boxes transmit the torque in the transmission variant shown in Figure 23 to two countershafts. This enables narrower gear widths despite the very high torques in the truck and a smaller center distance between the shafts to be chosen.

Mounted downstream of the main box is a range-change box. The transmission step between the two gears of the range-change box corresponds to the ratio spacing of the main box from first to fourth gear. This means that when shifting the range-change box gear from slow to fast gear at the same time the main box is shifted from fourth to first gear and vice versa. In this way, the spacing of the main box is utilized twice.

In this way, multiplying two splitter gears by four gears of the main box and two gears of the range-change box produces the maximum number of gears for these transmissions of 16 possible forward gears.

Synchromesh

During the gear change in the main box the movable gears are coupled with the shaft gears via shift dogs. The rotational speeds of the gears must be matched in the process. This task is performed by the synchromesh in manual transmissions for passenger cars. In a truck transmission the components of the synchromesh mechanism are not in the main box for weight reasons. In a manually shifted transmission it is therefore necessary to double-declutch when downshifting and double-clutch when upshifting. In this way the rotational speed of the transmission is adapted to the gear to be selected.

Accordingly, gearshifts for a manually shifted transmission on the one hand are uncomfortable and on the other hand need time for the driving speed to decrease again. For this reason automated manual transmissions have become prevalent in the truck sector. The rotational speed for the gearshift is adapted by means of the electronic transmission control.

Automatic planetary transmissions

The longitudinally or transversally mounted rear engine has gained acceptance in buses in order to guarantee maximum transport space for passengers and luggage.

The automatic planetary transmissions used predominantly in buses as a rule have six gears. These transmissions have essentially the same design as those in a passenger car, but their dimensions are adapted to the higher torque and mileage demands.

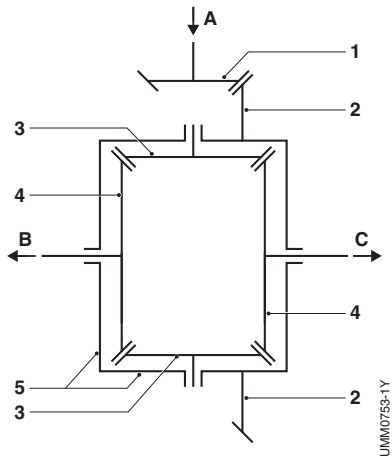
References

[1] H. Naunheimer; B. Bertsche; J. Ryborz; W. Novak; P. Fietkan: Fahrzeuggetriebe – Grundlagen, Auswahl, Auslegung und Konstruktion. 3. Aufl., Springer-Verlag, 2019.
[2] P. Gerigk; D. Bruhn; J. Göbert: Kraftfahrzeugmechatronik – Nutzfahrzeugtechnik, 1. Aufl., Westermann Schulbuchverlag, 2016.

Differentials

A differential has one input and two outputs. Differentials are most commonly used as final-drive differentials. Their purpose is to distribute uniformly the torque to the two drive wheels when driving straight ahead and when cornering. In all-wheel drive vehicles differentials are used as transfer cases to distribute torque between the vehicle axles.

Figure 1: Schematic diagram of a bevel-gear differential
A Input, propeller shaft,
B Output, drive shaft, left,
C Output, drive shaft, right.
1 Bevel-gear pinion,
2 Bevel crown wheel,
3 Differential bevel gears,
4 Drive-shaft gears,
5 Differential case.



Final-drive differentials

The final-drive differential distributes the torque coming from the manual transmission uniformly to the two drive shafts of a vehicle axle. Since when cornering the inner cornering wheel describes a smaller circle and thus turns more slowly than the outer cornering wheel, the final-drive differential must be able to compensate differential speeds between the drive wheels on an axle.

Bevel-gear differential
On a standard drive the input and output of the bevel-gear differential are offset by 90°. The bevel-gear differential drive transmits the torque from the bevel-gear pinion to the bevel crown wheel (Figure 1 and Figure 2). The bevel crown wheel (differential drive gear) transmits the torque via the differential case to the axle shafts of the differential bevel gears. The differential bevel gears distributes the torque to the two drive-shaft gears.

When the vehicle is driving straight ahead the two drive-shaft gears rotate at the same speed. The differential bevel gears circulate at the speed of the bevel crown wheel without rotating about their own axle shaft. They act as drivers and distribute the torque uniformly to the two drive shafts.

When cornering the differential bevel gears compensate the difference in speed of the drive shafts by rolling off on the drive-shaft gears rotating at different speeds. In this process the differential bevel gears rotate on their axle shafts, which are mounted in the differential case.

In most final-drive units the bevel-gear pinion meshes outside the axis center of the bevel crown wheel (Figure 3). This is called a hypoid drive. The advantages of a hypoid drive are:
– Smoother running since a greater number of gears mesh with each other.
– Higher torque transmittability since the bevel-gear pinion has a larger effective diameter due to the axial offset.

The drawback of this arrangement is the increase in sliding as opposed to rolling-off

between the bevel-gear pinion and the bevel crown wheel. However, this can be successfully countered with particularly pressure- and shear-resistant hypoid gear oils.

In front-transverse transmissions the bevel-gear pinion and the bevel crown wheel are replaced by two spur gears acting as pinion and differential drive gear (see Automated manual transmissions). Since the differential gears are designed as bevel gears, this design type is still classed as a bevel-gear differential.

Spur-gear differential

In a spur-gear differential the differential gears are designed as spur gears. There are various design types which utilize the different connection forms of planetary transmissions. In the differential pictured in Figure 4 the torque is routed via the differential drive gear directly into the planetary-gear carrier. The planetary-gear carrier distributes the torque to the planetary gears arranged in pairs on both sides. The planetary gears of a pair mesh with each other in each case with half the face width. The outer halves of the planetary gears mesh with the sun gears. The sun gears direct the torque in each case to the drive shafts [1]. Unlike a standard planetary transmission, there is no ring gear here. The differential case is the planetary-gear carrier for both planetary-gear sets. Sun gear 1 is at the same time drive-shaft gear 1, sun gear 2 is drive-shaft gear 2. Planetary gears 2 in Figure 4 are drawn with dashed lines because they are offset to the rear in the leaf plane and interlocked with planetary gears 1 with half the face width.

When the vehicle is driving straight ahead the two drive shafts rotate at the

Figure 3: Hypoid drive

1 Bevel-gear pinion,
2 Bevel crown wheel,
3 Axial offset.

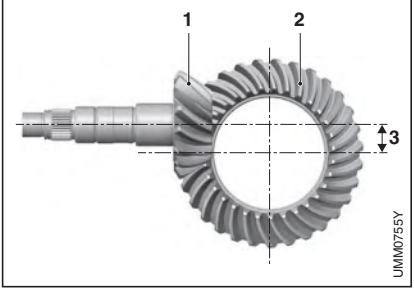
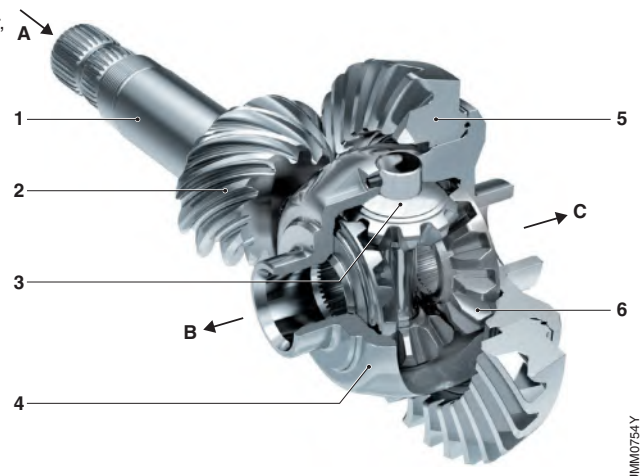


Figure 2: Structural design of a bevel-gear differential

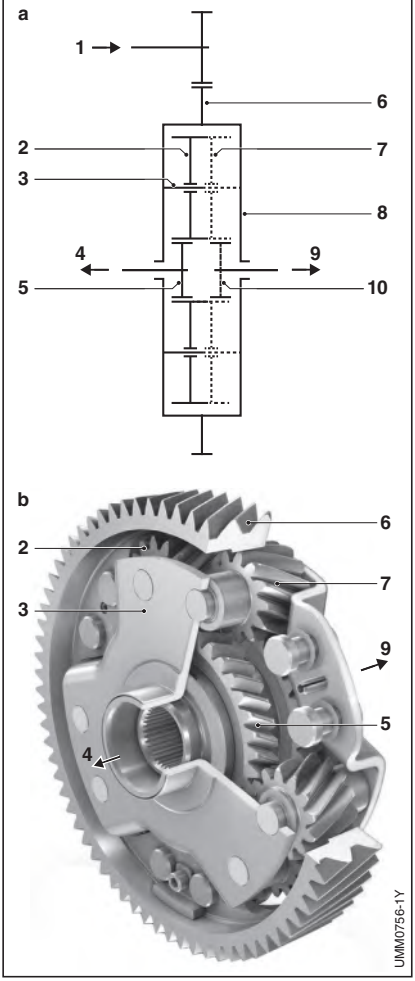
1 Drive shaft, bevel-gear pinion,
2 Bevel-gear pinion,
3 Differential bevel gear,
4 Differential case,
5 Bevel crown wheel,
6 Drive-shaft gear.
A Input, propeller shaft,
B Output,
drive shaft, left,
C Output,
drive shaft, right.



same speed. The planetary gears circulate at the speed of the differential drive gear without rotating about their own axle shaft. They act as drivers and distribute the torque uniformly to the two drive-shaft gears.

When cornering the planetary gears compensate the difference in speed of the drive shafts by rolling off on the sun gears rotating at different speeds. In this process the planetary gears rotate on their axle shafts in the planetary-gear carrier.

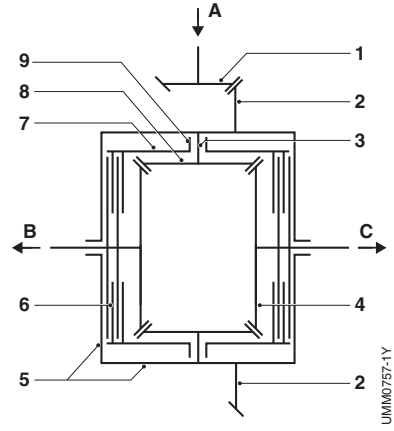
Figure 4: Spur-gear differential
a) Schematic diagram,
b) Structural design.
1 Input, 2 Planetary gear 1,
3 Planetary-gear carrier,
4 Output, drive shaft, left,
5 Sun gear 1, 6 Differential drive gear,
7 Planetary gear 2, 8 Differential case,
9 Output, drive shaft, right, 10 Sun gear 2.



Locking differential

A differential can transmit torque only as long as both vehicle wheels can transmit this torque to the roadway. As soon as the adhesion at a wheel breaks down (e.g. due to black ice), the wheel begins to spin. Since rotational speed but not torque is transmitted on the drive shaft of this wheel, the differential gears cannot be supported on the drive-shaft gear of the spinning wheel and torque is transmitted to the other drive-shaft gear. Consequently, the other wheel stops and transmits no drive torque, although adhesion with the roadway would be possible at this wheel in some cases.

Figure 5: Schematic diagram of a multi-disk self-locking differential
1 Bevel-gear pinion, 2 Bevel crown wheel,
3 Axle shaft, differential bevel gear,
4 Drive-shaft gear, 5 Differential case,
6 Multi-disk clutch,
7 Thrust rings, 8 Differential bevel gear,
9 Edge, thrust ring.
A Input, propeller shaft,
B Output, drive shaft, left,
C Output, drive shaft, right.



The purpose of the locking differential is to transmit drive torque to the two drive shafts, even if a wheel of a driver shaft has reduced adhesion with the roadway.

The locking differential has, in addition to the components described above, multi-disk clutches installed between the differential case and the driver shafts (Figure 5). One thrust ring per multi-disk clutch is installed in the differential case. The thrust rings are non-rotationally connected to the differential case but can move axially to be able to press against the multi-disk clutches. The axle shafts of the differential gears are mounted in recesses between the thrust rings (Figure 6). The axial forces that are produced during power transmission with bevel gears press the thrust rings onto the multi-disk clutches, and this is done more intensely the more torque is transmitted in the bevel-gear drive. The higher the transmitted torque, the higher the locking torque of the multi-disk clutch can be.

With equally good adhesion at both wheels, the majority of the torque is transmitted via the thrust rings to the differential gears, which distribute the torque – since they function as drivers – uniformly to the two drive shafts (Figure 5).

If, due to the loss of adhesion at a wheel, a difference in speed arises between the two drive shafts and thus also between the differential case and the drive shafts, torque is transmitted by the pressed multi-disk clutch to the drive shaft that is rotating at the slower speed.

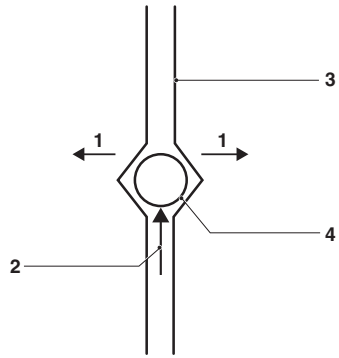
The level of torque and hence the degree of differential locking are dependent on the number of disks and their coefficient of friction, which can be varied by way of different friction linings.

Beyond this design variant there are further solutions approaches for locking differentials, e.g. an actively activatable multi-disk clutch between the drive shafts.

Figure 6: Schematic diagram of thrust rings and axle shafts of differential bevel gears of a multi-disk self-locking differential

The cutting plane is perpendicular to the cutting plane in Figure 5.

1 Spreading force, thrust rings, simultaneously contact pressure, multi-disk clutch,
2 Application of force from differential case,
3 Edge, thrust rings,
4 Axle shaft, differential bevel gear (top view).



Brake intervention with locking-differential effect

Since the introduction of antilock braking systems and driving-dynamics control (Electronic Stability Program) it has been possible to activate brakes of individual wheels. This opportunity is used to brake the faster wheel starting from a specific difference in speed between two drive shafts. Here, torque from the drive shaft rotating at the faster speed in relation to the body is supported via the brake. This enables the differential gears to transmit torque to the drive shaft rotating at the slower speed even without a lock in the differential.

References

[1] B.-R. Höhn; K. Michaelis; M. Heizenröther: Kompaktes Achsgetriebe für Fahrzeuge mit Frontantrieb und quer eingebautem Motor. ATZ – Automobiltechnische Zeitschrift Edition 1/2006.

Electronic transmission control

Functions and requirements

Hydraulics

Predominantly electronically actuated hydraulic systems are now used to control automated and automatic transmissions. Clutch actuators actuate dry clutches and gear selectors shift the gears in automated manually shifted transmissions and in dual-clutch transmissions. Shift elements activate the multi-disk clutches and multi-disk brakes in wet clutches and in automatic planetary transmissions.

The actuators, in this case the electrohydraulic actuators, serve as the connecting element between the electronics and the hydraulics. Actuators convert an electrical actuating signal output by the electronic control unit into an actuating force. Pressure-control valves ensure precise control of pressure levels at the shift elements (clutches and friction elements). As well as controlling, boosting, and distributing the hydraulic pressures and volumetric flows for the shift elements, this mechatronic system also supplies the torque converter, provides the lubricating pressure, and cools the automatic transmission.

Electromechanics

As an alternative to the hydraulic systems there are automated manually shifted transmissions and some dual-clutch transmissions with electromechanical control in which electric motors take on the role of the hydraulic system (actuate clutch, engage and select gear). These are designed for dynamic and wear reasons as BLDC (Brushless Direct Current) motors.

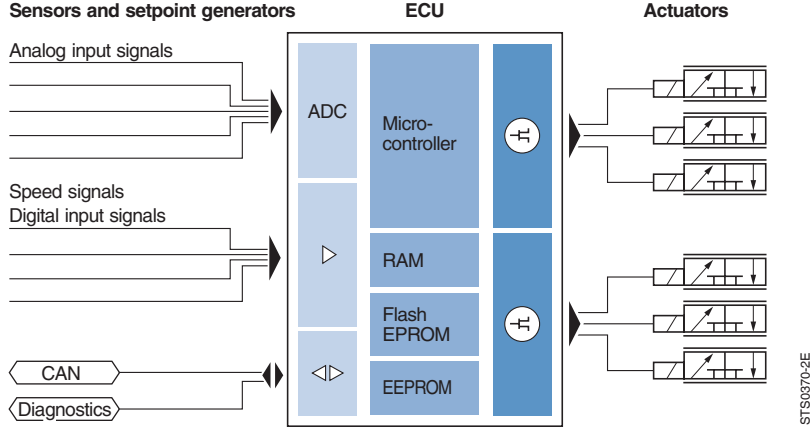
Electronics

Hydraulic actuation is retained for the clutches, while the electronics assume responsibility for gear selection, for gearshifting, and for adapting the pressures in accordance with the torque flow. The advantages are:

– a number of gearshift programs including adaptive drive functions,
– improved, constant gearshift comfort over the full service life,
– adaptability to different types of vehicle,
– simplified hydraulic control.

To guarantee these processes safely, comfortably and to suit the situation the control system requires various operating data which are acquired by sensors [1] (Figure 1):

Figure 1: Electronic transmission control



– transmission output-shaft speed,
– transmission input-shaft speed,
– pressure at the clutches,
– selector-lever position,
– positions of the program-selector and kick-down switches.

Further information such as engine speed, driver command, and engine torque is acquired by the engine ECU and transmitted via the CAN bus to the transmission control. The transmission control unit processes this information according to a predefined program and uses the results in determining the control variables to be output to the electrohydraulic actuators.

The transmission control must satisfy different requirements depending on transmission types and thus also varies greatly in its circuit complexity. Furthermore, the ambient conditions of different installation locations differ greatly in terms of temperature range, vibration, and tightness requirements. Implementations range from the simple remote-mounted ECU (the ECU is separated from the transmission) through an add-on ECU (e.g., for a dual-clutch transmission or a continuously variable transmission) to the complex electronic module for a dual-clutch transmission or automatic transmission which is installed inside the transmission housing (Table 2).

Transmission-control components

Components on the transmission

The transmission control comprises the following components:
– pressure-control and flow valves,
– hydraulic components (e.g., restrictors, dampers, slides, filters),
– control block (generally acts as the base carrier for the control unit),
– electronic transmission control unit.

The electrohydraulic control is accommodated in the transmission housing. The control is subject to further requirements with regard to temperature, change of temperature, contamination, and vibrations; it must also exhibit a high degree of media resistance to transmission oil [2].

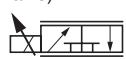
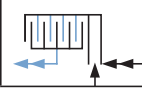
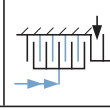
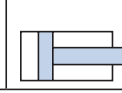
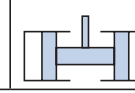
The conceptual distinction between actuator and shift element is shown in Table 1.

Table 2: Requirements relating to transmission control for different transmission types

	AT	CVT	DCT
No. of actuators	≤ 10	≤6	≤ 12
Pressure range [bar]	≤ 20 4…6	≤70 4...6	≤ 30 8…12
No. of sensors	High	Medium	Very high
Computing power			

AT Automatic Transmission
CVT Continuously Variable Transmission
DCT Dual Clutch Transmission

Table 1: Definition of actuator and shift element

Designation	Actuator	Shift element			
Installation location	Transmission control	Transmission			
Example	Electrohydraulic actuator (e.g., solenoid valve) 	Clutch 	Brake 	Clutch actuator 	Gear selector 

Option of intervention by the driver
Selector lever
The classic selector lever in the center console with the positions P, R, N and D and a cable to the transmission is replaced in today's shift-by-wire system by simple rotary or rocker switches which permit more freedom in the vehicle's passenger-compartment design. However, the main functions have been retained and are the same for all automated transmissions:
– P "Park"; the parking lock is activated in the transmission, preventing the vehicle from rolling,
– R "Reverse" engages the reverse gear,
– N "Neutral" or idle; i.e., there is no power transmission between the engine and transmission,
– D "Drive"; forward driving is activated.

Many systems still have a so-called M gutter which allows the gear to be manually selected.

Program selector
The driver can select the following programs by pushbutton or switch:
– E "Economy", results in economical driving in as high a gear as possible (Economy mode),
– S "Sport program", performance-oriented gear selection (Sport mode),
– M "Manual selection" of gear.

The Economy mode or Sport mode can also be selected by intelligent control in the software depending on the driver's driving performance.

Shift-point control

Drive programs
In selecting the gear to be engaged, the system refers to the rotating speed of the transmission output shaft and the driver command before triggering the appropriate solenoid valves or pressure regulators. The driver chooses between different drive programs (e.g., for optimum consumption- or performance-based driving) via the selector lever or an adaptive drive program is implemented which selects the optimum gear from the driving situation and driver type. In addition, the selector lever allows manual input from the driver.

Gearshift sequence from the example of an upshift in a multi-speed transmission
Power upshifts take place in an automatic transmission, unlike a manual transmission, without interruption of tractive force as overlap control. While clutch 1 opens for gear x, clutch 2 must close for gear y (Figure 2).

Figure 2: Overlap control
p_1 Pressure at cutting-in clutch,
p_2 Pressure at cutting-out clutch,
n_M Engine speed,
M Torque (total transmitted torque),
M_1 Torque at cutting-in clutch,
M_2 Torque at cutting-out clutch.

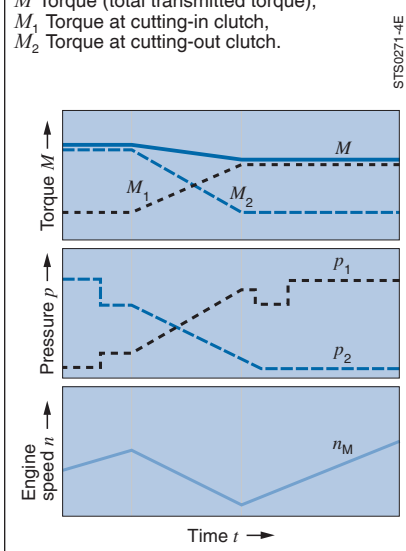
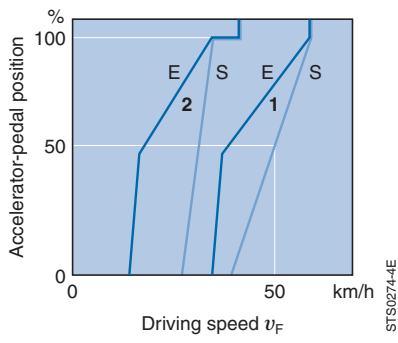


Figure 3 shows the upshift curve for shifting from first to second gear in Economy and in Sport modes, i.e., the accelerator-pedal position and speed at which the gear shift takes place. The same applies to downshifting. There must be a gap between the upshift and downshift curves (hysteresis) to avoid gear hunting.

Driving-situation detection
Variables such as linear and lateral acceleration are used for driving-situation detection, resulting in the detection of hills or corners. For this purpose progressive systems additionally incorporate the information from the navigation system in order to prepare future driving situations. The speed with which the accelerator pedal is pressed and the number of kick-down operations in a defined period of time or the frequency of brake-pedal operation serve to identify the type of driver (from very economy- to very performance-minded). A complex control program selects the appropriate gear for the current operating conditions and driving style. In other words, it suppresses trailing-throttle upshifts on the approach to and during corners, or automatically responds to slow throttle openings by activating a drive program for low engine-speed upshifts.

Concepts which combine the high level of convenience of such "intelligent" gearshift programs with facilities for active adaptation to individual driver preferences have become very widespread. The selector levers of these systems have a second parallel gutter (M gutter), in which simply touching the selector lever initiates an immediate gear shift (provided no engine-speed limits are exceeded). Alternative solutions come in the form of paddles or ± buttons on the steering wheel.

Figure 3: Upshift curve from first to second gear and downshift curve from second to first gear in Economy and Sport modes
1 Upshift curves,
2 Downshift curves.
E Economy mode,
S Sport mode,
Range above 100 % is triggered by kick-down switch.



Converter lockup
A converter lockup can be used in both multi-speed and continuously variable transmissions. A mechanical lockup clutch can be employed to improve the efficiency of the transmission unit by eliminating torque-converter slip. The variables employed to determine when conditions are suitable for activation of the converter lockup mechanism are engine load, transmission output-shaft speed, and state of the transmission.

Control of shift quality
The accuracy with which the pressure at the friction elements of the clutches is adjusted to the level of torque being transmitted (determined with reference to engine load and speed) has a decisive influence on shift quality; this pressure is regulated by a pressure regulator. Shifting comfort can be further enhanced by briefly reducing engine torque (via an engine intervention) for the duration of the shift (e.g., in a gasoline engine by retarding the ignition timing, reducing the injected fuel quantity, or cylinder cutout); this also reduces friction loss at the clutches and extends their service life.

Adaptive pressure control
Crucial parameters such as clutch fill time and slip time are permanently monitored to ensure a constant gearshift quality throughout the transmission's service life. In the event of deviations from specified limits, pressure corrections are made in the system in order to maintain an optimum time characteristic during the next gearshift and thus optimum shift quality.

Safety strategy
Special monitoring circuits and functions prevent transmission damage stemming from operator error, while the system responds to malfunctions in the electrical system by reverting to a backup mode. It is important to ensure at all times that vehicle instability does not arise.

Software-development requirements
Software development for automated transmissions must satisfy the following requirements:
– Compliance with functional safety (ASIL D, Automotive Safety Integrity Level, [3]),
– ASPICE (Automotive Software Performance Improvement and Capability Determination) as the development process in the international development network [4],
– AUTOSAR as the interface standard (Automotive Open Systems Architecture, [5]), and
– Software sharing between ECU manufacturers, transmission manufacturers, and vehicle manufacturers.

Hydraulic control

Functions
The main purpose of hydraulic control is to regulate, boost and distribute hydraulic pressures and volumetric flows. This includes generating the clutch pressures, supplying the torque converter, and providing the lubricating pressure. The hydraulic-control housings (hydraulic control plate) are made of diecast aluminum and contain a number of finish-machined slide valves and electrohydraulic actuators.

Figure 4 shows the electronic module with the sensors of electronic transmission control for a hydraulically actuated dual-clutch transmission. The electronic module is screwed to the hydraulic control plate and is seated inside the transmission wholly or partly in the transmission oil pan. The module is subject to high tightness and heat-resistance requirements in order to protect the electronics.

Electrically driven auxiliary oil pumps are increasingly used to generate the hydraulic pressure in line with demand, increase efficiency, and maintain the hydraulic pressure when the internal-combustion engine is stopped (e.g. in start/stop or hybrid systems). The activation and power electronics for the BLDC motor (Brushless Direct Current) are implemented either on the pump itself or as part of the electronic module.

Electrohydraulic actuators

Electrohydraulic actuators (electrohydraulic converter elements such as solenoid valves and pressure regulators) form the interface between the electronics with electrical signal processing and the hydraulics. They convert the low-power actuating signals into an actuating force with the increased power required for the process. The actuators monitor the oil flow and the pressure characteristics in the hydraulic control plate for this purpose. Actuators are consequently essential final control elements and form the basis of electrohydraulic transmission control. A solenoid and a hydraulic valve together form an electrohydraulic actuator.

General actuator requirements
Because they are installed directly in the transmission, electrohydraulic actuators are subject to the following specific requirements:
– endurance strength up to 5,000 hours or 300,000 km,
– vibration resistance up to 5,000 hours or 300,000 km,
– resistance to heat and insensitivity to change of temperature in the range of –40 to 160 °C,
– small size (saving on installation space and costs),
– media resistance to automatic transmission fluid with additives and water content,
– vibration insensitivity up to 30 g,
– generation of low hydraulic vibrations,
– operating time (100 %).

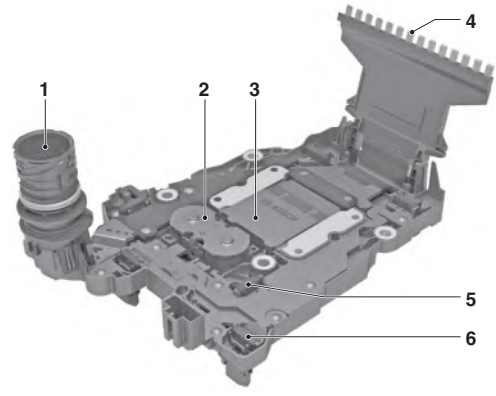
There are further specific requirements with regard to regulable pressure and throughflow as well as hysteresis and leakage.

Actuator design variants
There are different actuators, depending on function and design:
– switching valves (on-off solenoid valves, On/Off),
– pulse-width-modulated solenoid valves (PWM),
– flat-seat pressure-control valves (DR-F),
– spool pressure-control valves (DR-S),
– proportional solenoids (PM).

Solenoid valves
The on-off or PWM solenoid valves produce two different methods of converting an input signal (current or voltage) into an output signal (pressure).

Switching valve
The logic with the states 0 and 1 is sufficient for a simple switching valve and is generally controlled via the current (Fig-

Figure 4: Electronic module of a transmission control
1 Transmission connector (to wiring harness),
2 Pressure sensors,
3 ECU integrated in the electronic module,
4 Connectors for contacting the hydraulic valves,
5 Sensor domes of position sensors (rear side),
6 Sensor domes of speed sensors (rear side).



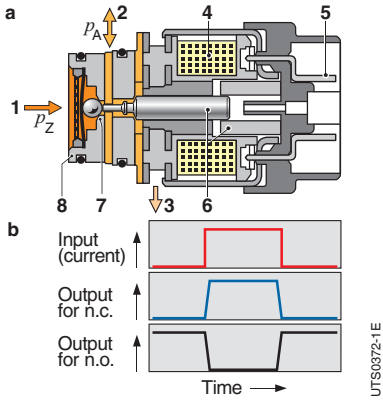
ure 5). This design is called on-off solenoid valve, the associated hydraulic circuit is fashioned as either a normally closed (n.c.) contact (closed at zero current) or a normally open (n.o.) contact (open at zero current).

The supply pressure p_Z generated by the transmission oil pump is applied ahead of the flange and closes the ball seat when the valve is closed at zero current. Because there is no pressure applied in the operating-pressure channel at zero current, the valve is "closed at zero current". In this state the operating pressure that supplies the consumers (e.g., a clutch) is connected directly to the return to the transmission-oil tank (oil pan) so that an applied pressure can decrease there or a contained oil volume can drain.

When a current is applied to the switching-valve winding the magnetic force draws the armature with the plunger connected to it towards the ball and opens the ball seat. The oil flows to the consumer and builds up the pump pressure there. At the same time the return to the transmission-oil tank closes.

These solenoid valves are characterized by low costs, low leakages, simple activation electronics, and low susceptibility to dirt. However, they are used increasingly less owing to the rising complexity and comfort requirements of transmissions [6].

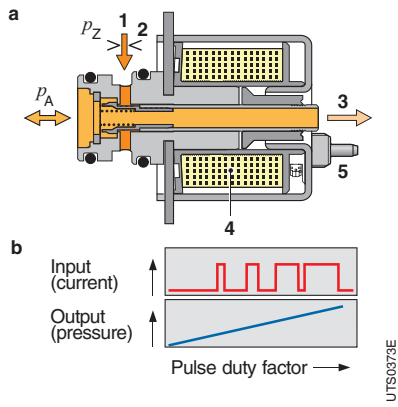
Figure 5: Switching valve (on-off solenoid valve)
a) Cutaway view of a valve closed at zero current,
b) Activation and characteristic curves.
1 Supply from transmission oil pump,
2 Operating-pressure channel (port A),
3 Return to transmission-oil tank,
4 Solenoid coil,
5 Electrical connection,
6 Armature with plunger,
7 Ball seat,
8 Flange (port P).
p_Z Supply pressure,
p_A Operating pressure.
n.c. normally closed (closed at zero current),
n.o. normally open (open at zero current).



PWM valves

Essentially, PWM valves are designed in exactly the same way as on-off solenoid valves. There are however designed for a higher switching speed and higher mechanical loads. To adjust pressures infinitely, they are used as so-called proportional valves. With the pulse-width-modulated input signal a volumetric flow proportional (rising) or inversely proportional (falling) to the pulse duty factor is generated which in turn generates the

Figure 6: Pulse-width-modulated solenoid valve
a) Cutaway view of valve,
b) Activation and characteristic curve.
1 Supply from transmission oil pump,
2 Preliminary restrictor,
3 Return to transmission-oil tank,
4 Solenoid coil,
5 Electrical connection.
p_Z Supply pressure,
p_A Operating pressure.



output pressure via the pressure drop at a preliminary restrictor (Figure 6).

PWM valves are characterized by properties such as low costs, low susceptibility to dirt, simple activation electronics, and a small hysteresis. However, PWM valves are used increasingly less because they generate undesirable noises.

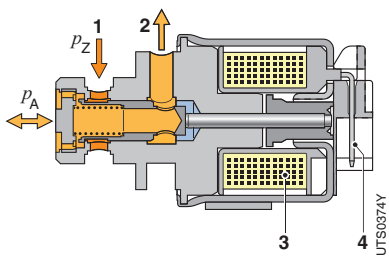
For important characteristic data, please refer to Table 3 [6].

PWM valve with high throughflow

The PWM valve with high throughflow (Figure 7) essentially has a similar design to the previously described standard PWM valve, but provides larger opening cross-sections with a larger diameter and longer opening lift of the closing element. This design requires a larger copper winding with a higher magnetic force.

Figure 7: Pulse-width-modulated solenoid valve for high throughflow

1 Supply from transmission oil pump,
2 Return to transmission-oil tank,
3 Solenoid coil,
4 Electrical connection.
p_Z Supply pressure,
p_A Operating pressure.



Pressure-control valve, flat-seat and spool designs

Two different pressure-control principles are used for analog valves: flat-seat and spool pressure-control valves [7]. Both principles involve complete closed-loop control circuits because the controlled pressure intervenes directly in the equilibrium of forces.

Flat-seat pressure-control valve

A flat-seat pressure-control valve (DR-F) is a proportional valve with a variable helix and a fixed helix (Figure 8). The DR-F op-

Figure 8: Flat-seat pressure-control valve (DR-F)

a) Schematic layout,
b) Falling characteristic curve.
1 Transmission oil pump,
2 Variable helix,
3 Restrictor (fixed helix),
4 Return to transmission-oil tank,
5 Plunger, 6 Solenoid coil,
7 Spring, 8 Electrical connection,
9 Control pressure, 10 Throughflow.
p_Z Supply pressure,
p_A Operating pressure.

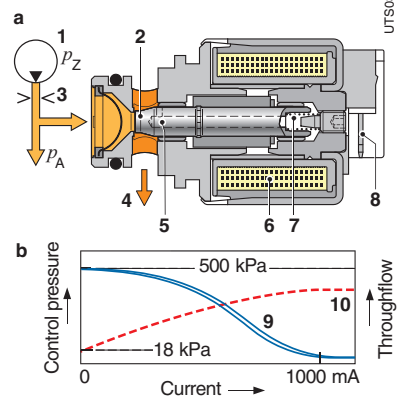


Table 3: Important characteristic data of on-off and PWM solenoid valves

Criterion	On-off valve	PWM valve	
Type		Falling curve	with high throughflow
Supply pressure [kPa]	400…600	300…600	400…1 200
Throughflow at 550 kPa [*l*/min]	>2.5	>1.5	>3.9
Clock frequency [Hz]	20	40…50	40…50
Coil resistance [Ω]	12.5	10	10

erates with an upstream restrictor (fixed helix) which is integrated either in the external hydraulic control or in the pressure regulator. The restrictor has a significant influence of the pressure-regulator characteristic, the precise matching of which is made possible by the integration of the restrictor in the pressure regulator. The ratio of the hydraulic resistances of the two restrictors cannot be arbitrarily small; a DR-F therefore has a residual pressure. This is why the effective control range cannot decrease to 0 kPa. This also results in a permanent pressure-regulator leakage. Energy losses are generated by this permanent oil flow into the transmission-oil tank. To compensate for this, transmission oil pumps with a higher volumetric flow are required, which is detrimental to the overall energy efficiency of a transmission and also results in higher fuel consumption.

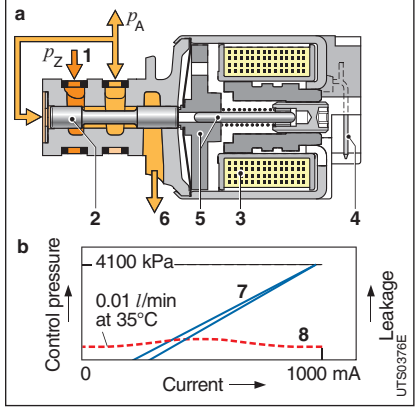
In a DR-F the hydraulic pressure (p_H), the magnetic force (F_M), and the spring force (F_F) influence the equilibrium of forces. The hydraulic pressure acts on the sensing surface of the piston. The magnetic force is proportional to the current that flows through the coil and a defined initial state is achieved with the aid of the spring force. When the flat seat (variable helix) opens oil flows back to the transmission-oil tank, bringing about a pressure reduction at the restrictor (fixed helix) and enabling the control pressure (p_A) to be adjusted. In the case of a flat-seat pressure-control valve with falling characteristic curve the magnetic force acts in the direction of the hydraulic pressure force (F_H) against the spring force. As a result the pressure in the control space decreases as current increases. In the case of a flat-seat pressure-control valve with rising characteristic curve the hydraulic pressure force and the spring force are in opposition to the magnetic force. Accordingly, the pressure in the control loop increases as current increases (Figure 8). A flat-seat pressure-control valve is accordingly an adjustable pressure-relief valve with hydraulic control function whose advantages disadvantages are set out in Table 4.

Spool pressure-control valve

A spool pressure-control valve (DR-S) operates with two helices and is also called a dual-helix pressure regulator or spool. The controlled pressure (p_A) is tapped between the inlet helix and the outlet helix (Figure 9). This results from the opening ratio at the inlet and outlet helices. Maximum pressure is obtained when the outlet helix is closed and zero pressure obtained when the inlet helix is closed. As in the flat-seat pressure-control valve the magnetic force, which is proportional to the current, the spring force, and the hydraulic pressure force, acts on the equilibrium of forces. The return of the controlled pressure to the sensing surface, e.g., the end face of the spool, via an internal channel closes the control loop. The control loop can also be closed via oil channels in the control plate. A spool pressure-control valve can like a flat-seat pressure-control valve be designed with a rising or falling characteristic curve.

Figure 9: Spool pressure-control valve (DR-S)

a) Schematic layout,
b) Rising curve.
1 Supply from transmission oil pump,
2 Spool, 3 Solenoid coil,
4 Electrical connection,
5 Armature with plunger,
6 Return to transmission-oil tank,
7 Control pressure, 8 Leakage.
p_Z Supply pressure,
p_A Operating pressure.



The advantages and disadvantages of a spool pressure-control valve are set out in Table 5.

Proportional solenoid
Inserting a separating plane is between the hydraulic spool and the solenoid section of the spool pressure-control valve produces a proportional (linear-force) solenoid (PM) and a spool. The advantage of this separation is the use of standardized proportional solenoids and functional adaptation by individual spools. Otherwise they are functionally identical to spool pressure-control valves.

Table 4: Advantages and disadvantages of a flat-seat pressure-control valve (DR-F)

Advantages	Disadvantages
High accuracy	High leakage
Low cost	Residual pressure
Not susceptible to dirt	Costly electronics
Insensitive to disturbances	Difficult damping
	Seat rebound

Table 5: Advantages and disadvantages of a spool pressure-control valve (DR-S)

Advantages	Disadvantages
High accuracy	High susceptibility to dirt
Insensitive to disturbances	Costly manufacture of precision parts
Low temperature sensitivity	Costly electronics
Low leakage	
No stops	
Dampable	
Zero pressure achievable	

Characteristic values
Typical characteristic values of flat-seat and spool pressure-control valves are set out in Table 6.

Use of different actuators
Table 7 shows the respective links between the individual transmission functions and the actuator types that can be used in a multi-speed transmission and in a continuously variable transmission (CVT). The decision whether a PWM valve, on-off valve, DR-F, DR-S or PM is used also depends on among others the following system factors [7]:
– installation location,
– dynamic properties of the system,
– leakage,
– accuracy,
– costs,
– dirt concentration and dirt composition.

Not to be underestimated either are the experiences of the customer with the respective actuator types (e.g., risk, confidence level) or even the effort to be applied in switching existing control concepts to other actuators.

Table 6: Typical characteristic values for flat-seat (DR-F) and spool pressure-control valves

Criterion	DR-SI	DR-F	
Type	DR-S	D26 falling curve	D20 rising curve
Supply pressure [kPa]	700…2500	500…600	500…800
Controlled typical pressure [kPa]	0…2500	40…600	40…800
Typical current range [mA]	0…1200	150…770	150…1000
Dither frequency [Hz]	<200	<200	<200
Chopper frequency [kHz]	3	1	1
Diameter [mm]	28	26	23

Methods for controlling shift elements
There are two different pressure-control variations to initiate the actuating process of a shift element (clutch, brake or gear selector): direct control and pilot control [8].

Pilot control
A large volumetric flow is required to enable a shift element (e.g., a clutch) to be quickly filled. To control large volumetric flows the actuators must likewise be of such a very large size to be able to deliver the necessary forces. To avoid this, pilot control operates with two control circuits, the pilot control circuit (0 to 6 bar) and the main control circuit (approx. 20 bar). The electrohydraulic actuator is situated in the pilot control circuit. A lower pressure is applied at this actuator (pilot control valve) than at the check and booster valves, which are located in the main control circuit. The pilot control valve controls these spool valves by making hydraulic fluid available at the respective control surfaces (Figure 10).

Table 7: Linking of transmission functions and actuator types

Multi-speed transmission					
Function	PWM (old)	On/Off	DR-F	DR-S	PM
Control main pressure	×		×	×	×
Trigger gear change			×	×	×
Modulate shifting pressure			×	×	×
Control converter lockup clutch	×		×	×	×
Reverse-gear interlock		×			×
Safety functions (fail-safe)		×	×		×

Continuously variable transmission (CVT)					
Function	PWM (old)	On/Off	DR-F	DR-S	PM
Adjust gear ratio			×	×	×
Control belt tension			×	×	×
Control starting clutch			×	×	×
Reverse-gear interlock		×			×

Direct control
In the case of direct control the electrohydraulic actuator is integrated directly in the main control circuit. As a result, there is no need for the supply, check, and booster valves used in pilot control. The volume provided by the pump and the pressure made available to actuate the shift element are thus controlled directly by the electrohydraulic actuator (Figure 11).

Figure 10: Hydraulic layout of pilot control

1 Transmission oil pump,
2 Main pressure valve,
3 Electrohydraulic actuator,
4 Supply valve,
5 Check valve,
6 Shift element.
7 Booster valve.

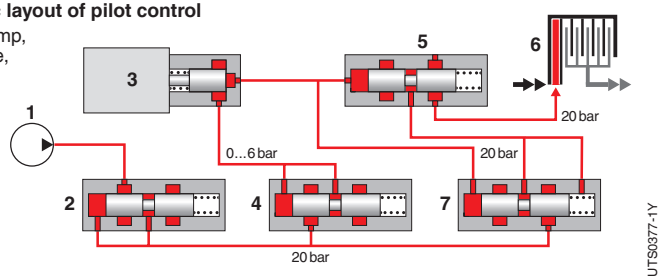
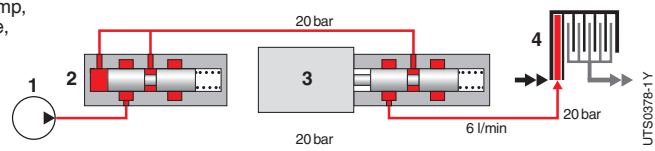


Figure 11: Hydraulic layout of direct control

1 Transmission oil pump,
2 Main pressure valve,
3 Electrohydraulic actuator,
4 Shift element.



References

[1] W. Schmid: GS-Seminar Getriebesysteme – Aktuatorik zur Steuerung von Kfz-Automatikgetrieben. Robert Bosch GmbH, Department GS-TC/ENS, Presentation Documents, Schwieberdingen, 2008.

[2] K. Reif: Aktoren und Module zur Getriebesteuerung. Vieweg+Teubner Verlag, Wiesbaden, 2010.

[3] https://www.all-electronics.de/abkuerzungsverzeichnis/asil/.

[4] https://www.lhpes.com/blog/what-is-aspice-in-automotive.

[5] http://autosar.org/.

[6] Robert Bosch GmbH – GS-TC/ENS-P1: Produkteinführung GS. Presentation Documents, Schwieberdingen, 2010.

[7] Robert Bosch GmbH: Elektronische Getriebesteuerung. 2004.

[8] Robert Bosch GmbH: Produkteinführung GS. Department GS-TC/ENS-P1, Presentation Documents, Schwieberdingen, 2010.

Internal-combustion engines

Thermal engines

Operating principles and concept

Internal-combustion engines are classified as thermal engines. The essential feature of a thermal engine is cycle direction, which is characterized by the output of work.

Contrasting with thermal engines are heat pumps, also called refrigerating machines, which are characterized by cycle direction in the opposite direction and require drive power in order to operate.

The operating principle of thermal engines is always the same. A working medium is compressed, whereupon energy is supplied while the medium is compressed with a corresponding further increase in pressure. This is followed by an expansion with power output. In open cycles the working medium which has performed the work is discharged. In closed cycles the initial state must be re-established by cooling the working medium before compression is restarted in open and closed cycles.

Table 1: Characteristics and functioning principles of important thermal engines

Thermal engine	Steam circuit	Stirling	Steam engine	Gas turbine	Jet engine	Reciprocating piston engine	Wankel rotary engine
Thermodynamic reference cycle	Rankine	Ericson	Steam cycle	Joule	Joule	Seiliger (Otto/ Diesel)	Seiliger
Typical working medium	H_2O, ethanol	Air, helium	H_2O	Air	Air	Air, air/ fuel mixture	Air, air/ fuel mixture
Cycle direction	Closed		Open				
Energy input	From outside by heat transfer			From inside			
Thermodynamic energy input	Stationary	Non-stationary	Non-stationary	Stationary	Stationary	Non-stationary	Non-stationary
Typical energy sources, fuels	Coal, fuels, uranium	Any heat source	Coal, fuels	Methane, ethane, propane, butane	Kerosine	Gasoline, Diesel fuel,	Gasoline
Transmission of work performed	Turbine	Reciprocating piston, rotary piston (rotor)	Piston	Turbine	Linear momentum	Reciprocating piston	Rotary piston (rotor)
Typical maximum pressure	50 bar	3 bar (air)	50 bar	40 bar	40 bar	200 bar	60 bar
Typical minimum pressure	0.05 bar	1 bar	1 bar	1 bar	1 bar	1 bar	1 bar
Typical max. efficiency	40 %	30 %	~25 %	40 %	40 %	~42 % / ~45 %	~30 to 35 %
Operating concept	Compression, Heating, Evaporation, Superheating, Expansion, Condensing	Compression, Heating, Expansion, Cooling	Compression, Heating, Evaporation, Superheating, Expansion, Exhaust	Induction, Compression, Combustion, Expansion		Induction, Compression, Combustion, Expansion, Exhaust	

Many thermal engines are characterized by an energy input resulting from a combustion process (Table 1). During combustion the energy chemically bound in the fuel is supplied as reaction heat to the cycle. Here compounds containing carbon and hydrogen oxidize with oxygen, which is why typically ambient air with approximately 21 % volume of oxygen makes up a significant part of the working medium.

Cycle direction

The factor that is crucial to cycle direction is the energy input. Here, a distinction is initially made between stationary (or continuous) energy input and non-stationary (or cyclical) energy input. What all piston engines including the Stirling engine have in common is a non-stationary energy input which only ever occurs close to compression dead center with minimum cylinder volume.

A particular feature of all open cycles is an internal energy input which is achieved by the addition and combustion of fuel. In contrast, closed cycles require an energy input via heat exchangers. Here there is no direct contact between the working medium and combustion processes, with the exception of heat conduction. Unique in this context is the steam engine, which vaporizes the working medium by means of a heat flow from an external source and then supplies it to the piston engine.

Another particular feature of the various thermal engines is the different energy sources which are used. Three different energy sources can be used: solids, liquids and gases. The main advantage of thermal engines which operate with an open cycle and an internal energy input is that they do not require any heat exchangers for cycle direction and are therefore compact in design. This advantage can be amplified by using liquid fuels with a high energy density. Gas engines, for example when deployed in passenger cars and commercial vehicles, are also becoming more attractive (running costs, good fuel consumption). The internal-combustion engine can thus be derived from a selection of thermal engines.

Efficiency

The internal-combustion engine is characterized by open cycle direction with internal combustion. Non-stationary operation facilitates the intake and compression of the working medium at mass-averaged peak temperatures above 2,500 K and averaged peak pressures above 200 bar with a very good maximum efficiency above 40 %.

Stationary cycles, being limited by the properties of the materials, do not achieve mass-averaged peak pressures and temperatures of this magnitude, but instead achieve maximum local peak temperatures around 2,500 K. Gas turbines therefore operate at a lower level of efficiency. The closed cycle of the steam circuit achieves a higher level of efficiency than gas turbines with a likewise moderate pressure of around 50 bar. This is achieved by a significant lowering of the low-pressure level. The other thermal engines are much lower in terms of their maximum efficiency.

Reciprocating-piston engine

The internal-combustion engine, in its reciprocating-piston-engine variation, is the primary thermal engines used for mobile applications. A wide range of fuels is conceivable in principle, but diesel and gasoline are still the main energy sources used.

Real cycles

A cycle is a thermodynamic process which has identical initial and final states (see Thermodynamics). The cycle usually passes through several changes of state in such a way that with a thermal engine work is drawn from the cycle. Here the working medium of the cycle undergoes thermodynamic changes of state.

Ideal reference cycles are suitable for demonstrating basic correlations. In the case of new, as yet unknown engines, they help to provide an overview of their method of operation and efficiencies. However, a real cycle calculation is required for detailed analyses.

Real cycles are set in proportion to the ideal cycles. This means that the heat capacities are considered for example depending on temperature or pressure.

The chemically altered composition of the smoke gases is also approximated in the substance variables in order to take into consideration the changed physical properties based on combustion. In particular, what is assumed is not an adiabatic change of state, but rather at least one polytropic curve with an exponent adapted to the heat loss, or even the heat loss to the wall is used, e.g. with the Woschni approach from the Reynolds similarity theory [1].

The charge cycle is calculated with its dissipation losses (flow losses, real flow cross-sections, see Fluid mechanics) and thus the residual exhaust gas is also taken into account. Furthermore, empirical approaches are normally used for friction and the fuel calorific value is calculated depending on the air ratio. Ultimately, in particular the heat addition (combustion process, heating process) and the heat dissipation (heat transfer) are modeled in detail.

One way to quickly assess real cycles is to describe them using an efficiency chain. Here the real cycles are successively depicted by taking individual characteristic quantities into account.

Overall efficiency
The overall efficiency or effective efficiency η_{eff} sets the effectively available power P_{eff} in proportion to the energy flow $\dot{Q}_{\text{add}} = \dot{m}_{\text{B}} H_{\text{u}}$ which was added by the fuel mass flow $\dot{m}_{\text{B}}$ and its lower calorific value H_{u}:

$$\eta_{\text{eff}} = \frac{P_{\text{eff}}}{\dot{Q}_{\text{add}}} .$$

Diesel engines have at higher loads an effective efficiency of approx. 45 %, large slow-running diesel engines have a much higher effective efficiency. Gasoline engines have, depending on the combustion process, an effective efficiency at their best point of approx. 42 %.

Mechanical efficiency
The mechanical efficiency sets the effectively measured power P_{eff} in proportion to the pressure-indicated cycle power P_{ind}. The indicated power is determined from the work W – this is the area of the real pressure characteristic against volume $\int pdV$ – and the time t per working cycle according to the following relationship:

$$P_{\text{ind}} = \frac{dW}{dt} \approx \frac{\Delta W}{\Delta t} .$$

The effective power differs from the indicated power essentially in the frictional losses (piston, bearings), the power-transmission losses from control elements (camshaft, valves), and the power of the accessories (oil and water pumps, fuel-injection pump, alternator). For the mechanical efficiency:

$$\eta_{\text{m}} = \frac{P_{\text{eff}}}{P_{\text{ind}}} .$$

Normal mechanical efficiencies are load-dependent and at full load come in at just under 90 %, while at low part load (10 % load) values around 70 % are registered.

Efficiency-of-cycle factor
The efficiency-of-cycle factor describes the efficiency with which the real cycle can be approximated by the chosen reference cycle. It therefore contains losses resulting in particular from dissipation losses. For a detailed loss analysis it is recommended to split the efficiency-of-cycle factor between the high-pressure loop and the charge-cycle loop (Table 2).

Normally, for calculation purposes, an ideal gas with temperature-dependent heat capacities is assumed and the cycle is used with geometrically identical dimensions, the same air ratio, without residual exhaust gas, complete combustion, and thermally isolated walls. An engine described in this way is also called a “perfect engine”. The efficiency-of-cycle factor at full load is in the range of approximately 80 to 90 %.

If an ideal gas with constant heat capacity is still to be used, it is possible to introduce the “efficiency of the perfect engine”, which sets the power of the “perfect engine” in proportion to the power of the “ideal cycle”.

Fuel-conversion factor
In particular, gasoline engines with rich-mixture combustion (air/fuel ratio $\lambda < 1$) are subject to high HC and CO emissions, which usually cannot be taken into account in the approach of added heat via the fuel calorific value H_{uB}. However, the exothermy of these gases H_u is considerable, which is usually also reflected in higher exhaust-gas temperatures after the oxidation-type catalytic converter. It is taken into account in the fuel-conversion factor:

$$\eta_B = \frac{(H_{uB} - H_u)}{H_{uB}} .$$

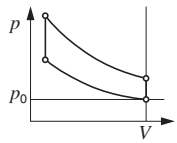
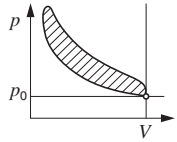
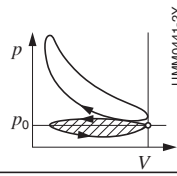
$\eta_B = 1$ is usually set for diesel engines. In gasoline engines the figure can drop to 0.95, and with very rich air ratios $\lambda < 1$ even further below that.

Efficiency chain
The entire efficiency chain can be described as follows (Table 2):

$$\eta_{eff} = \eta_i \, \eta_m = \eta_{th} \, \eta_g \, \eta_m .$$

References
[1] G. Woschni, Die Berechnung der Wandverluste und der thermischen Belastung der Bauteile von Dieselmotoren, MTZ 31 (1970).

Table 2. Graphic representations and definitions of the individual and overall efficiencies of the reciprocating-piston engine
The hatched areas indicate the newly added work component of the characteristic efficiency quantity. The efficiencies are explained in the text.

Pressure vs. volume diagram	Designation	Boundary conditions	Definition	Efficiencies
	Theoretical reference cycle, e.g. constant-volume cycle	Ideal gas, constant specific heat, infinitely rapid heat addition and dissipation, etc.	$\eta_{th} = 1 - \varepsilon^{1-\kappa}$ Theoretical or thermal efficiency	η_{th}
	Real high-pressure working cycle	Wall heat losses, real gas, finitely rapid heat addition and dissipation, variable specific heat	η_{gHD} Efficiency factor of the high-pressure cycle	η_g
	Real charge cycle (4-stroke)	Flow losses, heating of the mixture or the air, etc.	η_{gLW} Gas-exchange efficiency	
Mechanical losses, cannot be suitably represented in the pressure vs. volume diagram	Losses due to friction, cooling, accessories	Real engine	η_m	η_m η_m

η_i η_{eff}

Mixture formation, combustion, emissions

All internal-combustion engines share the common feature of combustion occurring after induction of the fresh mixture or air and after subsequent compression. In reciprocating-piston engines with internal combustion this takes place close to top dead center (TDC). The result is a pressure increase, which is transmitted via the piston and the connecting rod to the crankshaft in the form of a crankshaft torque (Figure 1).

The sequence of compression and subsequent combustion on the one hand has a significant impact on the pressure characteristic and thus on the efficiency and the torque output. On the other hand, this sequence defines the creation of emissions inside the engine. In this respect, gasoline and diesel engines differ in their process control.

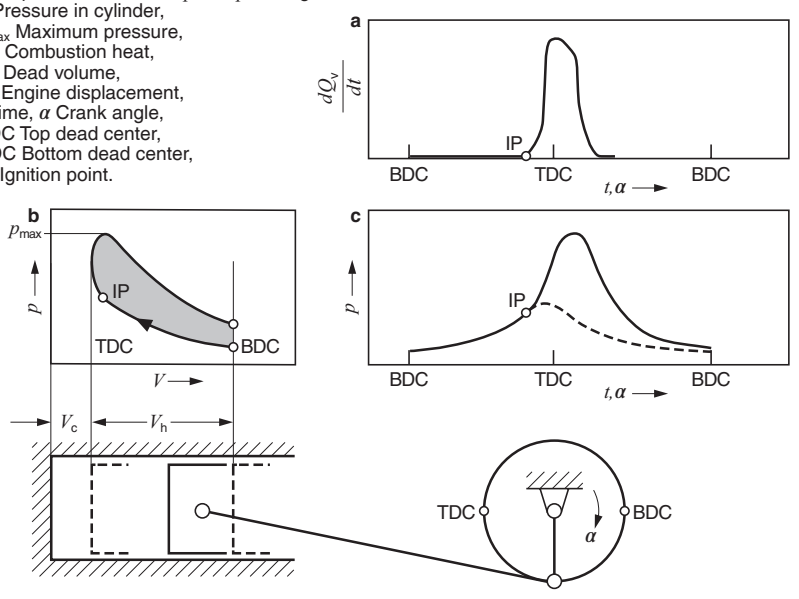
Gasoline engine

The characteristic feature of a gasoline engine is that it uses an external ignition source, normally an electrode spark plug. Ideally a suitable homogeneous air/fuel mixture is created to provide the required flammability. This is achieved by means of external mixture formation (intake-manifold injection) or internal mixture formation (gasoline direct injection).

Mixture formation

For the most part, a homogeneous mixture preparation is effected in a gasoline engine, i.e. the intake air is fully mixed with the vaporized or atomized fuel during the induction and compression strokes. The excellent vaporization qualities of gasoline enable it to be injected into the intake manifold. Modern stratified-charge combustion processes are, on the other hand,

Figure 1: The engine power cycle
a) Course of combustion,
b) Representation in the p-V diagram (enclosed area corresponds to indicated work),
c) Representation in the p-t or p-α diagram,
p Pressure in cylinder,
p_{max} Maximum pressure,
Q_V Combustion heat,
V_c Dead volume,
V_h Engine displacement,
t Time, α Crank angle,
TDC Top dead center,
BDC Bottom dead center,
IP Ignition point.



characterized by partially heterogeneous mixture formation (see Stratified-charge operation).

Mixture formation is decisively influenced by the vaporization conditions, the injection pressure, the cylinder-charge movement, and the time available with the aim of homogenization. Essentially, mixture formation involves the interaction of two processes: droplet vaporization caused by the temperature difference (Figure 2) and droplet disintegration by aerodynamic forces (Figure 3). Intake-manifold injection and direct injection differ here (Table 1).

Intake-manifold injection

In the case of intake-manifold injection, a film of mixture is created on the intake-manifold wall ahead of the intake valve whose fuel mass is reduced more intensively as the air velocity increases. This air velocity changes along broadly linear lines with the engine speed. On

Figure 2: Fuel vaporization
1 Diesel, 2 Petroleum, 3 Gasoline.

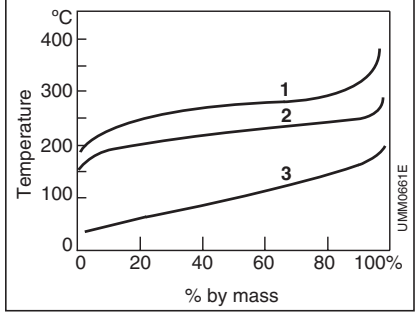
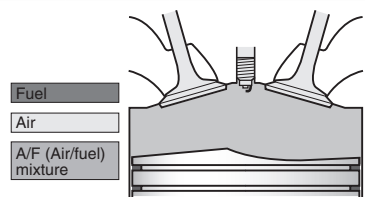
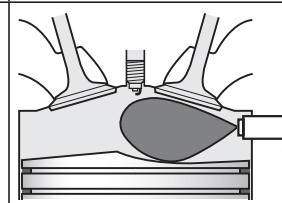


Table 1: Operating strategies of gasoline engines

Operating strategy	Stoichio-metric	Rich	Lean	Ultra-lean	Stratified-charge operation lean
Mixture formation	Homogeneous				Heterogeneous and homogeneous
Combustion-chamber composition					
Fuel injection	Intake-manifold injection and direct injection				Direct injection
Ignition	External ignition source	External ignition source	External ignition source	Auto-ignition	External ignition source
Typical compression ratio	8…12		11…13	12…16	11…14
Load control	Quantity				Quality
Operating range	Entire program map	Full load, high speed range	Entire program map	Part load	Part load
Application, development stage	Conventional, series		Gas engines, series	Research stage	New combustion processes

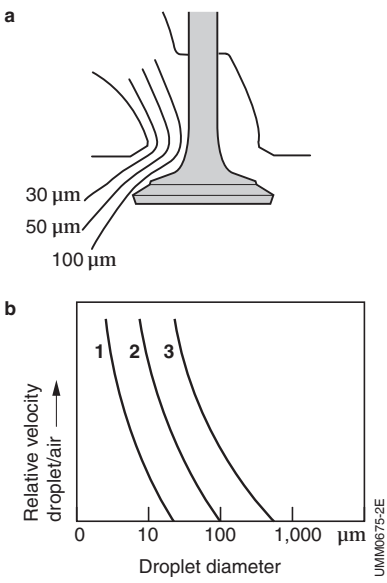
account of the low temperature and incomplete vaporization in the intake manifold with the resulting formation of film, intake-manifold injection occurs at a very low injection-pressure level below 10 bar.

The dynamics of the wall-applied fuel film and the mechanisms involved during vaporization are one of the main causes of inaccurate fuel metering, above all during transient engine operation. Only the smaller droplets entrained with the intake flow reach the inside of the cylinder (Figure 3). Typically, their characteristic diameter is already less than 30 µm. Here, droplet acceleration is proportional to the ratio of relative velocity to the air and to the droplet diameter.

The very high turbulence intensity and the high flow velocities in the valve gap result in very good mixture formation.

Figure 3: Droplet formation

a) Characteristic flow situation in valve gap,
b) Interaction between droplet diameter and relative velocity of droplets with regard to air flow.
1 0 %,
2 50 %,
3 Over 90 %.



As process control progresses the remaining small fuel droplets assume the temperature of the mixture and vaporize (Figure 4). This enables the vaporized fuel to distribute more effectively in the combustion chamber, resulting in homogenization. An optimum combustion-chamber design prevents intensive contact of fuel with the wall, since this always involves the risk of fuel condensation.

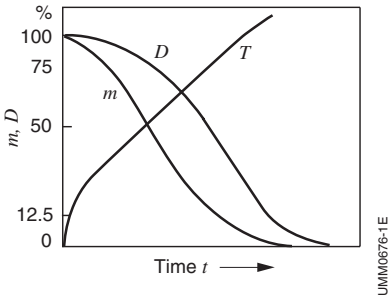
Gasoline direct injection

Gasoline direct injection does not utilize the mechanisms of mixture formation in the valve gap. A higher injection pressure of 50 to 350 bar is therefore required. Injection is completed by no later than the gas-flow bottom dead center – i.e. at the end of the induction stroke – to allow sufficient time for homogenization during the compression stroke. Multiple injections are also carried out in the induction and compression strokes in order to achieve better homogenization. This increases the charge movement, which in turn encourages mixing. In heterogeneous combustion processes injection occurs only at the end of the compression stroke (see Stratified-charge operation).

The mixture included in the combustion chamber is now compressed, depending first and foremost on the throttle-valve position and the compression ratio, to

Figure 4: Droplet vaporization

m Mass, D Diameter, and T Temperature of a vaporizing fuel droplet.
m and D are shown scaled to the maximum value. No scale is specified for T, because only the basic characteristic is involved.



a pressure level of 10 to 40 bar. This equates to a temperature level of 300 to 500 °C, depending first and foremost on the ambient temperature and the compression ratio.

The advantage of direct injection is that the fuel is metered exactly. Vaporization of the fuel in the combustion chamber also requires the cylinder charge to be adequately cooled. This enables the compression ratio to be increased by roughly one unit, resulting in greater efficiency.

Combustion in gasoline engines

In all combustion processes combustion – oxidation – occurs only at the end of the compression phase and in the early expansion phase. The subsequent combustion process differs depending on the nature of mixture formation (homogeneous or heterogeneous). An entirely homogeneous mixture reacts in a premixed combustion, an entirely heterogeneous mixture in a mixture-controlled combustion. During the stratified-charge operation of modern direct-injection engines, a large proportion (>50 %) of the injected fuel is likewise homogenized up to the start of combustion.

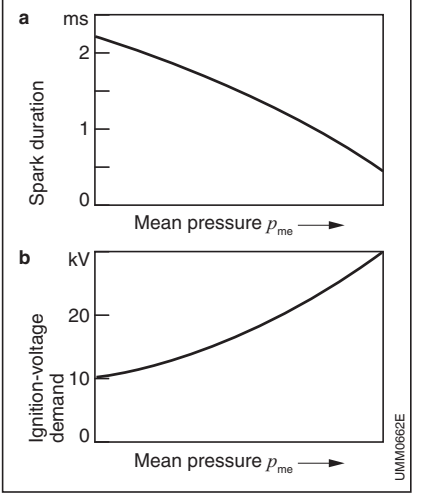
In both homogeneous and partly heterogeneous mixture formation, actual combustion is preceded by ignition and the ignition phase.

Ignition

Ignition is typically performed with the aid of an electrode spark plug. When a high voltage is applied, a sparkover occurs between the electrodes, depending on the mixture state (i.e. pressure, temperature, and mixture composition), the electrode gaps, and the air humidity. The high voltage is typically in the range of 10 to 40 kV. First and foremost the number of molecules between the electrodes influences the ignition-voltage demand (Paschen's law). The mixture ignited by the spark must in the course of its combustion release that amount of energy that is needed to ignite the immediately adjacent mixture. As leaning of the mixture increases, the energy content of this mixture decreases when the electrode gap remains constant. This is accompanied by an increase in the energy demand needed to ignite the adjacent – also lean – mixture. By enlarging the electrode gap, it is possible to increase the volume ignited by the spark and thereby raise the energy content. However, enlarging the electrode gap requires an increase in the ignition voltage. In this way, this increases for example in lean-combustion processes or in the event of a load increase. In the event of a load increase, the spark duration simultaneously decreases as the ignition-voltage demand increases (Figure 5). On account of heat losses at the spark-plug electrodes, heat-convection losses and cyclically fluctuating mixture states, the electrical ignition energy is above the theoretical minimum ignition energy by up to an order of magnitude (Figure 6). The stochastically fluctuating states (flow field and mixture state) between the electrodes are the main cause of large cyclic variations in a gasoline engine. This situation is improved by enlarging the electrode gap. Modern engines here already operate with a maximum value above 1 mm. Enlarging the electrode gap requires an increases in the ignition voltage.

Figure 5: Ignition sparks and ignition-voltage demand

a) Spark duration,
b) Ignition-voltage demand U_Z.
The ignition-voltage demand rises as exhaust-gas recirculation or leaning increases.



The purpose of ignition is to ignite the air/fuel mixture and thereby initiate the actual combustion process. Depending on the velocity of the subsequent combustion and on the piston speed (and thus on the engine speed), the point of ignition must be variably adapted (Figure 7).

Figure 6: Minimum ignition energy for a methane/propane mixture

1 Ignition energy for a static mixture,
2 Ignition energy for a mixture with a flow velocity of 6 m/s,
3 Ignition energy for a mixture with a flow velocity of 15 m/s.
With an increase in the proportion of inert gas the ignition-voltage demand increases for all the flow states shown.

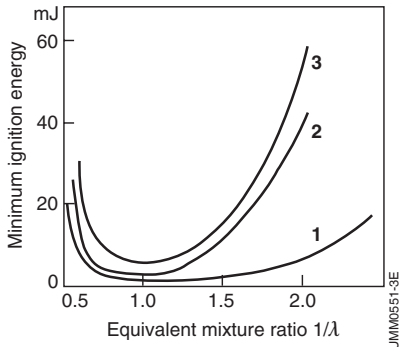
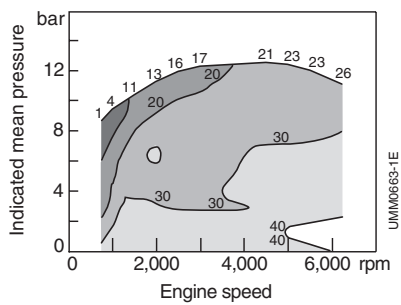


Figure 7: Program map of ignition point

Values in °CA before TDC for a homogeneously operated gasoline engine.
Indicated mean pressure is obtained from the indicated work/(engine speed times engine displacement).



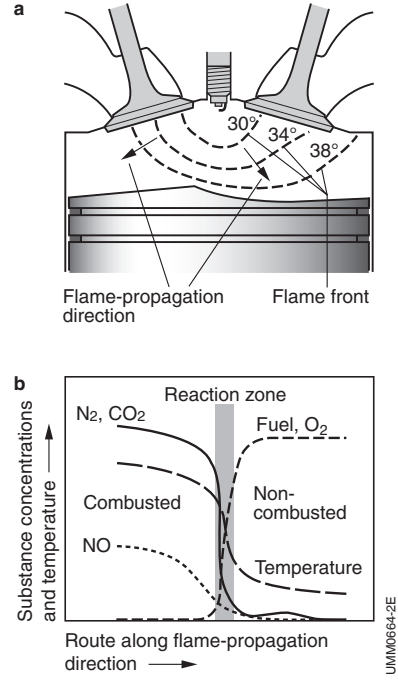
Combustion of homogeneous mixtures

Propagation of the flame front

Ignition sets in progress the actual combustion in a homogeneously operated gasoline engine. Here the flame propagates from the spark plug. It is possible to define a flame-front velocity which is composed of the sum of the flame velocity and the flame-front movement (charge movement, expansion caused by density differences). In a homogeneously operated gasoline engine a good distinction can be made between the combusted and non-combusted mixture, since the flame continuously propagates outwards (Figure 8).

Figure 8: Flame propagation and premixed combustion

a) Flame-front propagation in a gasoline engine during homogeneous operation (the figures in degrees are calculated from the ignition point),
b) Temperature and substance concentrations along the flame-propagation direction.



It takes a few milliseconds for the flame to reach a radius of around one centimeter in order to propagate unhindered from the piston recess and cylinder head at a velocity well in excess of 10 m/s.

Influences on flame velocity

The decisive factor now is the propagation velocity of the flame, which is also referred to as the turbulent flame velocity. The faster the flame propagates, the better the combustion inside the engine. The following factors encourage a high flame velocity: a low proportion of inert gas, a temperature increase in the non-combusted mixture, a pressure increase, and a high turbulence level.

Most fuels have their maximum flame velocity in slightly rich-burn operation in the range of λ = 0.85 to 0.9. A further advantage of slightly rich-burn engine operation is the cooling effect provided by the excess fuel. Racing engines and passenger-car engines in the rated-output range are therefore operated with slightly rich mixtures.

Increasing the proportion of inert gas reduces the flame velocity. A practical application of inert-gas variation is exhaust-gas recirculation (EGR), in which combusted exhaust gas with the chief constituents CO_2, H_2O and N_2 is added to the air/fuel mixture. As a rule of thumb an exhaust-gas recirculation rate of 10 % already reduces the flame velocity by 20 %.

The reason why modern gasoline engines can nevertheless be operated with high internal exhaust-gas recirculation rates is the influence of temperature. Doubling the temperature increases the flame velocity by a factor of roughly 4.

Cylinder pressure has a lesser influence; an increase in pressure causes a slight acceleration of the flame velocity.

Turbulences in the combustion chamber

The turbulence level in the combustion chamber has the greatest influence on the combustion velocity. The flame velocity varies along roughly linear lines with the turbulence intensity. Turbulence intensity is a measure of the high-frequency change in flow velocity at a given point in the combustion chamber [1]. The turbulent kinetic energy is proportional to the square of the turbulence intensity.

Turbulence intensity is a three-dimensional quantity which is influenced above all by the flow profile of the charge movement in the combustion chamber. The velocity of the flow processes inside the engine that increases linearly to the greatest possible extent as engine speed increases is extremely important. As the flow velocity increases, so too does the turbulence intensity in the combustion chamber. This is the reason for stable engine operation over a very wide engine-speed range. The flame would otherwise no longer be able to burn cleanly at increasing engine speed and consistent flame velocity because of the shorter time available for combustion. However, the positive contribution of turbulence cannot equal completely the influence of speed, which is why combustion at high speeds nevertheless stretches over a wider crankshaft range. This is an additional reason why gasoline engines operate with less efficiency at higher speeds.

In gasoline engines the turbulence in the combustion chamber is vitally important to energy conversion. The decisive factor in causing turbulence is the charge movement in the cylinder, which is fundamentally influenced by the intake flow (depending on the configuration of the ports in the cylinder head) and the combustion-chamber shape (Figure 9).

Combustion causes a pressure increase which can also be simultaneously heard. For comfort reasons, suitable combustion-adjustment measures must be taken to minimize this pressure increase as much as possible. This poses a conflict of objectives with thermodynamic efficiency. The maximum pressure-increase gradient of gasoline engines is in the range of between 0.5 and 3 bar/°CA.

Combustion of partially homogeneous mixtures

Modern stratified-charge combustion processes facilitate operation with excess air in the part-load range. Effective mean pressures of p_{me} < 1 bar make even operation with a mean excess-air factor λ > 5 possible in stratified-charge operation. The main benefit here is the improved charge cycle, because throttling (which has a negative impact on overall efficiency) can be largely dispensed with.

Conventional operation is not possible without throttling in the lower part-load range, because homogeneous mixtures leaned in this way burn much too slowly and therefore incompletely. The solution is to provide local fuel stratification in the area of the spark plug by means of an optimized injection strategy in the late compression phase. The difficulty here lies in optimally matching the injection strategy and ignition, since the conditions between the electrodes change (Figure 10).

Figure 9: Turbulent and mean kinetic energy as a function of the crankshaft position, each referred to mass

1 Turbulent kinetic energy,
2 Mean kinetic energy.

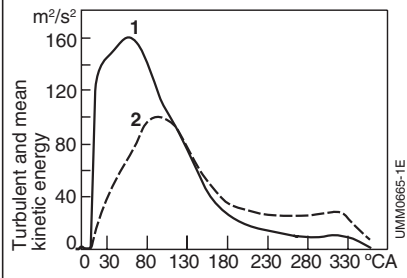


Figure 10: Excess-air factor λ over time in a spherical volume (with radius r) around the electrode midpoint

1 Stratified-charge combustion process, λ within r = 2 mm (gaseous fuel),
2 Stratified-charge combustion process, λ within r = 3 mm (gaseous fuel),
3 Stratified-charge combustion process, λ within r = 5 mm (gaseous fuel),
4 Stratified-charge combustion process, λ within r = 5 mm (liquid and gaseous fuel),
5 Optimized stratified-charge combustion process,
6 Homogeneous combustion process.

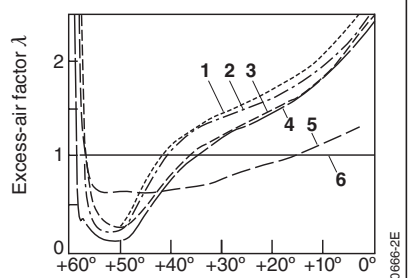
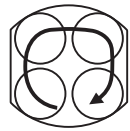
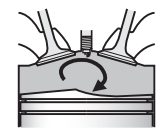
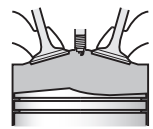
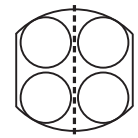
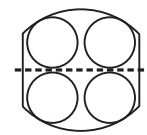


Table 2: Charge cycle and flow profile

Operating strategy	Swirl	Tumble	
Flow			
Location of flow axis	 Vertical axis z	 Lateral axis x	 Lateral axis y

UMM0660-1E

Charge movement
Large-scale, swirling and circular flows with diameters similar to the characteristic quantities of the combustion chamber provide the basis for the cylinder-charge movement. A basic distinction is made between flows about the vertical axis (cylinder axis), which are called swirl, and flows about the two lateral axes (crankshaft axis and perpendicular to this), which are called tumble (Table 2). In reality there is an overlap of the three basic flows, which gives rise to complex three-dimensional flow fields. Tumble and swirl essentially differ in their behavior inside the engine.

Tumble flow disintegrates up to the moment when compression dead center is reached and contributes primarily to flame propagation in the first combustion half.

Swirl flow lasts longer into the subsequent expansion phase. The disintegration of the large swirls into ever-decreasing, turbulent structures encourages the creation of turbulence. In the continuing process, however, the viscosity of the working medium results in a disintegration with a detrimental effect on the combustion velocity.

The creation of turbulence is assisted by the combustion-chamber geometry. In particular, flows in the piston-recess or squish-gap area (area between piston and cylinder head) support flame propagation (see e.g. [2]).

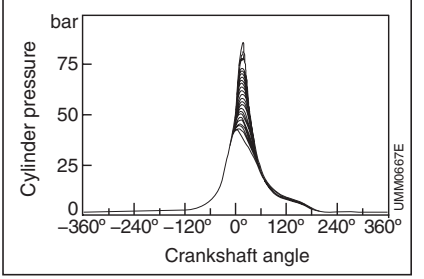
One of the greatest physical challenges associated with fashioning homogeneous combustion processes is the combustion behavior in the expansion phase, since typically over 10 % of the supplied fuel is still not converted 30° CA after TDC. At this point the non-combusted mixture is still only in the immediate vicinity of the wall and must still be converted after re-emerging from the top-land area between piston ring and cylinder wall (see e.g. [2]). This phenomenon results in incomplete combustion in the final phase.

Uncontrolled combustion
Numerous undesirable processes make it difficult to fashion a homogeneous combustion process. In addition to cyclic variations, knock and auto-ignition characteristics have detrimental effects. Extreme forms of advanced ignitions can occur, particularly in modern, highly turbocharged engines.

Cyclic variations
Cyclic variations are a consequence of the premixed combustion of homogeneous combustion processes which reacts sensitively to numerous disturbance values (Figure 11). Such disturbance values are, for example, the mixture composition, the residual-exhaust-gas content, the thermodynamic state variables, and the flow profile. All the values vary from working cycle to working cycle and cause noticeable cyclic fluctuations in energy conversion. In particular, leaning with excess air of more than 40 % increases the cyclic variations significantly. Initiation of ignition has the greatest influence on cyclic variations.

Knocking
Another challenge is posed by the undesirable phenomenon of knocking. Knocking occurs in response to increases in pressure and temperature after the onset of regular combustion until the local conditions stimulate auto-ignition. Here, the residual mixture in response to the temperature increase as a result of the

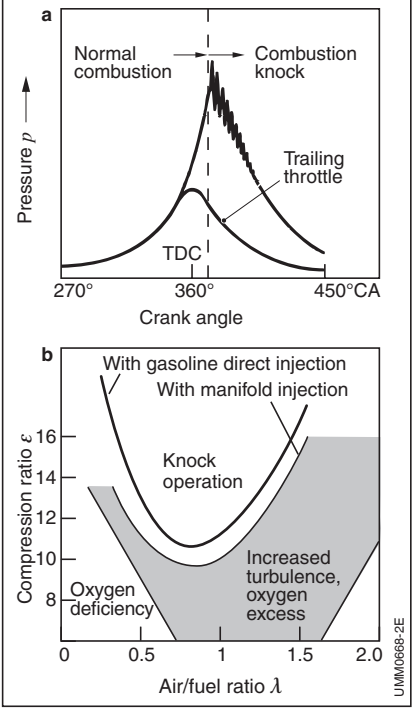
Figure 11: Cyclic fluctuations of the cylinder pressure of the individual working cycles during stoichiometric operation (λ = 1)



progressing combustion reaches the ignition temperature and burns almost simultaneously without a further controlled flame propagation. The pressure pulsations generated in the process result in material wear in the engine bearings and cause serious engine damage in the event of prolonged operation. The temperature peaks can also result in damage to components. Typically, knocking only then occurs in a working cycle if 80 % of the fuel in the combustion chamber has not been combusted. Knocking can be observed particularly at low engine speeds with suitably sufficient time for auto-ignition and at high loads with high combustion-chamber temperatures. Fuels with high ignition temperatures such as, for example, methane or ethane reduce knock sensitivity. Knocking combustion can be reduced by shifting the energy conversion in the retardation direction (ignition timing towards TDC). High-compression and turbocharged engines are more knock-sensitive because of their higher final compression temperatures (Figure 12). Useful measures to counter knocking include effective cylinder cooling of the hot areas, for example also by the vaporization effect of gasoline direct injection, increasing the turbulence, reducing the geometrical compression ratio, and optimizing the fuel, for example with additives.

Figure 12: Compression ratio, knocking and knocking combustion

a) Example of knocking combustion,
b) Effect of compression ratio and air/fuel ratio on the operating limits (knocking and engine misfires). Direct injection instead of intake-manifold injection and increased turbulence shift the operating limits.



Auto-ignitions

In contrast to knocking, auto-ignitions can occur quite perfectly even in very retarded compression situations. Possible causes of auto-ignitions are, for example:

– Over-retarded moment of ignition with incomplete combustion, thus fuel-film formation causing auto-ignitions,
– Full-load operation with high cylinder-component temperatures,
– Ignition caused by abrasion and hot particulates,
– Oil emissions caused by defective or worn piston rings.

Extreme forms of auto-ignition can occur in turbocharged and high-compression gasoline engines; these can result in peak pressures in excess of 150 bar and serious damage. However, these extreme auto-ignitions occur only very rarely with a probability of less than 0.01 per thousand.

Pollutant formation and pollutant reduction in gasoline engines

In addition to the unavoidable combustion products of carbon dioxide (CO_2) and water (H_2O), the concentration of which is dependent on the composition of the fuel, nitrogen oxides (NO_x), non-combusted hydrocarbons (HC) and carbon monoxide (CO) are the main emissions from gasoline engines (Figure 13).

Pollutants

Nitrogen oxides

Nitrogen oxides (NO_x) require four factors in order to be created: oxygen, nitrogen, high temperatures, and time. Because oxygen and nitrogen are determined by the mixture composition in gasoline engines and the available time is defined by way of the engine speed, nitrogen oxides in gasoline engines can only be reduced by low maximum combustion-chamber temperatures (e.g. by means of ignition retardation and exhaust-gas recirculation).

Hydrocarbons and carbon monoxide

Hydrocarbons (HC) and carbon-monoxide emissions (CO) are the result of incomplete combustion. In rich-burn operation there is, on account of the lack of oxygen, an increase in both HC and CO emissions. In lean-burn operation with accordingly reduced flame temperature a more intensive extinguishing of the flame is encountered, above all in the fire-land area close to the wall, with a resulting increase in HC emissions. Because of the excess oxygen the slightly oxidizable CO emissions are nevertheless reduced.

Soot and particulates

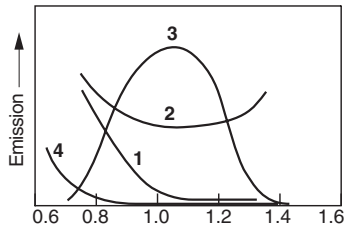
Engines with intake-manifold injection cause – primarily in extreme rich mode – increased soot emissions. These are however lower compared with gasoline engines with direct injection. With the Euro 6c emission-control standard the particulate limit for gasoline engines with Direct Injection has been reduced to 1/10 of the value applicable to Euro 6b such that a particulate filter must be used in many vehicle models.

Sulfur oxide

Sulfur-oxide emissions are dependent on the sulfur content in the fuel. Currently only a sulfur content of 10 mg/kg is permitted in Europe so that sulfur-oxide emissions are of secondary importance.

Thanks to exhaust-gas treatment, modern homogeneous gasoline engines are, once they have reached the catalytic-converter operating temperature, virtually zero-emission machines. Three-way catalytic converters in operation at $\lambda = 1$ reduce the nitrogen-oxide emissions while simultaneously oxidizing HC and CO molecules (see Three-way catalytic converter). Alternative approaches are required in lean-burn operation. For this reason, NO_x storage catalytic converters are typically used for stratified-charge concepts. These catalysts store the nitrogen oxides. Rich-burn engine operation applied at regular intervals reduces the stored nitrogen oxides at high temperatures. Because NO_x storage catalytic converters are sensitive to sulfur contamination, desulfation cycles must additionally be completed in slight rich-burn operation at temperatures in excess of 600 °C (see NO_x storage catalytic converter).

Figure 13: Gasoline-engine emissions
1 Carbon monoxide (CO),
2 Hydrocarbons (HC),
3 Nitrogen oxides (NO_x),
4 Soot.



Load control in gasoline engines

In homogeneously operated gasoline engines the load is adjusted by means of the injected fuel mass. The corresponding air mass is adapted by means of the throttle-valve position for the required operation at λ = 1. This is known as quantity control. In the part-load range this causes induction throttling, which is detrimental to the overall efficiency. The cause of this is the work done during induction and emission of the gas, the share of which must be subtracted from the energy generated in the combustion process. This disadvantage can be partially compensated for by varying the valve timing. Advanced or retarded "intake closes", reduced valve lift or retarded "exhaust closes", in which hot exhaust gas is additionally inducted, are typical measures. Alternatively, external exhaust gas can be recirculated (exhaust-gas recirculation) in order to reduce throttling. In turbocharged gasoline engines the air flow rate and thus the fuel-mass flow rate is adjusted in the upper load range typically by means of the position of the wastegates on the turbocharger (see Exhaust-gas turbochargers).

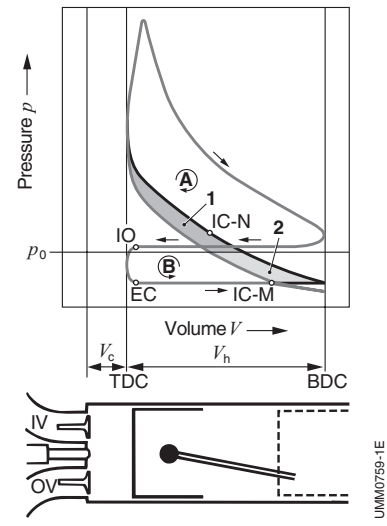
In engines with stratified charge the injected fuel mass is adjusted by the load in unthrottled operation at part load. This is known as quality control. The transition between stratified- and homogeneous-charge operation in the mid-load range requires control-engineering complexity and expenditure.

Power yield and efficiency

The part-load performance of a gasoline engine deteriorates as a result of charge-cycle losses (throttling), poor process control (peak pressures below 30 bar), and an increasing amount of engine friction. Because even at passenger-car driving speeds in excess of 100 km/h most vehicle engines still operate in the part-load range, measures to increase the efficiency in this operating range prove highly successful. Examples of such measures are:

Figure 14: Miller process: Power yield and efficiency increase by dethrottling via valve timing

1 Additional useful work,
2 Less charge-cycle losses.
TDC Top dead center,
BDC Bottom dead center,
IV Intake valve,
EV Exhaust valve,
IO Intake opens,
IC-N Intake closes (normal operation),
IC-M Intake closes (Miller process),
EC Exhaust closes,
V_c Compression volume,
V_h Cylinder displacement,
p_0 Atmospheric pressure (1013 mbar),
A High-pressure loop,
B Charge-cycle loop.



– Displacement reduction (downsizing),
– Cylinder shutoff (e.g. in V8 and V12 engines),
– Dethrottling (stratified charge, exhaust-gas recirculation, valve timing),
– Increased compression ratio,
– Longer transmission ratio to lower the engine-speed level.

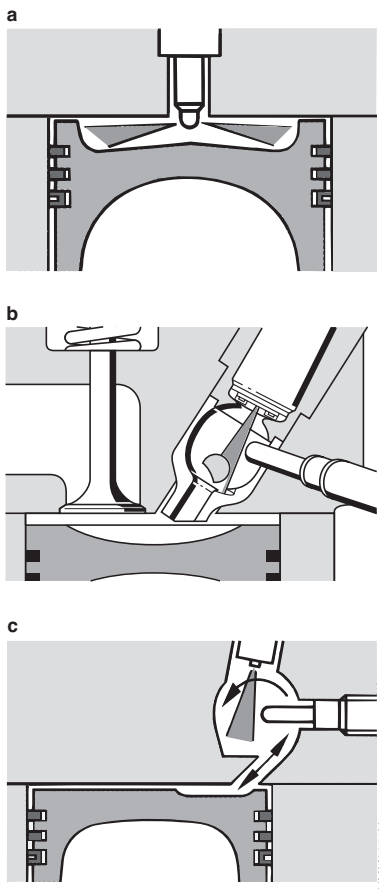
In the Miller process the intake valve is closed at a very advanced stage. In this way, the throttle valve can remain open and the throttling losses are lower (Figure 14).

Diesel engine

A significant characteristic feature of diesel engines is the absence of external ignition. This is achieved by injecting the ignitable fuels into highly compressed and therefore hot air. High final compression temperatures and pressures to levels in excess of 600 °C and 100 bar in turbocharged engines provide for extremely stable engine operation. Fuel-spray formation, fuel vaporization, fuel mixing, and subsequent combustion can take place within a very short period of time.

Combustion processes with direct injection have in the last few decades established themselves as the favored choice over indirect combustion processes such as whirl-chamber or prechamber systems. In indirect combustion processes the local flow in the secondary volume is essentially responsible for fuel preparation (Figure 15).

Figure 15: Diesel-engine combustion processes
Combustion-chamber shape and nozzle arrangement for:
a) Direct injection,
b) Prechamber system,
c) Whirl-chamber system.



Mixture formation

Mixture formation is dominated by the interaction of the injection spray with the flow field in the combustion chamber. Here, the challenge consists in quickly injecting and preparing relative large masses of fuel of up to 200 mg per liter displacement. A typical injection duration is in the range of 1 ms. The term used in relation to the fuel mass flow into the combustion chamber is the injection rate (unit: kg/s). This fuel mass flow is typically injected by multihole injection nozzles.

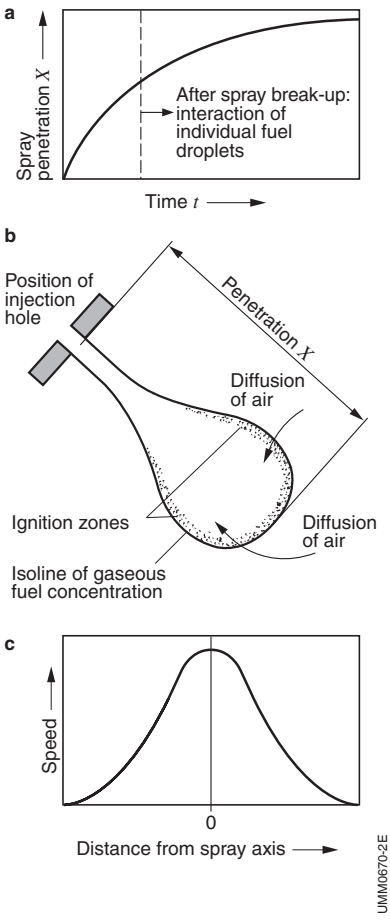
A combination of between four and ten injection holes with a diameter of between 120 and 150 µm is commonly used. As well as the small hole diameter, the high injection pressure of over 2000 bar encourages fast fuel injection and mixture formation.

The diameter of a characteristic spray is initially the same as that of the nozzle hole. After it has traveled a few millimeters, the spray breaks up into individual droplets which interact with the flow field. The liquid phase of the fuel spray can, especially depending on the density of the working medium (air, air-fuel mixture or air with EGR), penetrate a few centimeters into the combustion chamber before the spray is completely atomized or vaporized (Figure 16).

Turbulences encourage the formation of droplets and the vaporization of the fuel. In modern diesel engines more than 80 % of the turbulence intensity in the area of the fuel spray is generated by the injection. This is further assisted by the charge movement, whereby diesel engines with flat cylinder head experience predominantly more swirl than tumble

flows. An additional contribution can be made by compression-induced air flows from the outer to the inner area of the combustion chamber ("squish flows") or by the combustion-chamber design, e.g. by vaporization-inducing contact with the hot piston-recess area.

Figure 16: Spray propagation and mixture formation in diesel engines
a) Characteristic spray penetration behavior,
b) Typical injection-spray formation (the lean edge areas with infinite λ are outside the continuous line),
c) Velocity profile of injection spray.

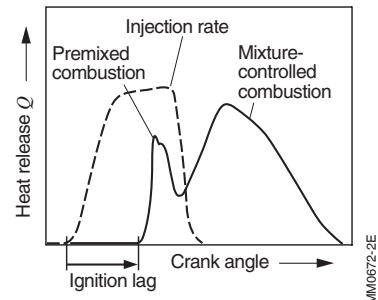


Combustion in diesel engines
Combustion in diesel engines differs from conventional gasoline-engine operation in the compression-ignition behavior. In all, diesel-engine combustion can always be described in three consecutive processes: ignition lag, premixed combustion, and mixture-controlled combustion. Depending on the operating state and the map range, the three processes have different time components (Figure 17).

Ignition lag
Ignition lag refers to the period of time between start of injection and start of combustion. It is crucially dependent on the cylinder temperature, the cylinder pressure, and the combustibility of the fuel. Both mixture formation and the first chemical pre-reactions of the air/fuel mixture take place in the ignition-lag phase. The ignition lag is long during cold engine operation or when a low-quality fuel with a low cetane number is used. The influence of cylinder pressure is less dominant than that of temperature. Basically, however, an increase in cylinder pressure reduces the ignition lag. The fuel injected during the ignition-lag phase does not burn yet. The ignition lag can be between 0.1 ms in the rated-output range to over 10 ms in the cold-start range.

Premixed combustion
The ignition lag defines the phase of premixed combustion by means of the injected and as yet non-combusted fuel. The longer the ignition lag, the more fuel

Figure 17: Diesel-engine combustion



is premixed in combusted form. This fuel mass can perfectly well be over 20 mg per liter displacement. Combustion typically starts at the edge of the fuel spray, where the fuel has mixed very well with air and therefore optimum ignition conditions are present in terms of temperature and excess-air factor λ. The exothermic reaction results in a local temperature increase with temperatures above 2300 K, which quickly converts the as yet non-combusted premixed fuel in a chain reaction. Here, the chemical reactions that take place determine the rate of combustion. The self-accelerating chain reaction results in an extremely fast combustion with large pressure-increase gradients. For this reason, the premixed converted fuel mass must be kept to a minimum in diesel engines. This is typically achieved by means of a pilot injection whose local combustion exerts an initial temperature increase with a reducing effect on the ignition lag of the subsequent main injection.

Mixture-controlled combustion

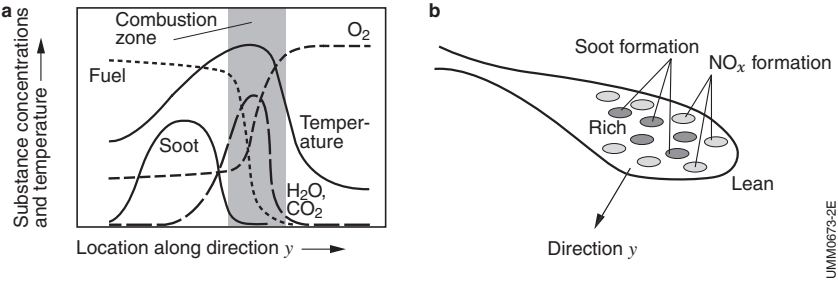
The amount of premixed converted fuel can be between less than 1 % in the upper full-load range and 100 % in the lowest part-load range. The remaining amount burns in a mixture-controlled combustion. In contrast to premixed combustion, in the case of mixture-controlled combustion – also called diffusion combustion – the transfer of oxygen into the combustion zone defines the conversion rate. It is difficult to distinguish between a combusted zone and a non-combusted zone, because there is no precisely defined flame front. Basically, the diffusion flame establishes itself at the edge of the spray and burns in a limited range at $0.8 < \lambda < 1.4$. As the boundary conditions change (e.g. further fuel vaporization, oxygen transfer, wall contact), so too the reaction zone strays to where locally stoichiometric conditions prevail (Figure 18).

Turbulences

Mixture-controlled combustion dominates in the high load range, in which large quantities of fuel are injected. Here, the mixture-formation and combustion processes occur in parallel. As with premixed combustion, the conversion rate can also be influenced by the injection. A lesser influence, but also of an accelerating character, is exerted by increases in both temperature and pressure and by a reduction in the proportion of inert gas. The dominant factors are mixture formation and the transfer of oxygen into the combustion zone by a high local turbulence.

For this reason, turbulence intensity is the deciding variable in fashioning diesel-engine combustion processes and is reflected in ever higher injection pressures with a high kinetic energy on the part of the fuel spray, which is subsequently converted into turbulent kinetic energy. Local turbulence causes the vital transfer of oxygen into the local reaction zones. The cylinder-charge movement (swirl, squish flows) supports this phenomenon, the main contribution however being made by the pulse of the injection spray. As well

Figure 18: Mixture-controlled combustion
a) Substance concentrations and temperature distribution,
b) Formation of pollutants inside the injected spray.



as increasing the injection pressure, increasing the hole diameter or the number of holes is also conceivable. However, an increase in the injection rate results for the most part in a local over-enrichment with a negative impact on fuel conversion.

Particular characteristics of diesel-engine combustion

Cold-start processes

Cold-start processes in diesel engines pose a particular challenge, particularly at outside temperatures below –10 °C. At starter speeds down to below 100 rpm a large proportion of the cylinder charge escapes through the piston rings during the relatively slow compression phase. In addition, the cold cylinder temperatures give rise to increased wall heat losses. Results experienced are low peak pressures below 30 bar and, depending on the outside temperature, low peak temperatures below 400 °C.

Fuel vaporization at top dead center

Fuel vaporization at top dead center causes further cooling. The outcome is very long ignition lags. In an extreme-case scenario there may be no ignition whatsoever with the result that fuel can accumulate in the cylinder on the cylinder wall and piston over several working cycles. Its ignition after a few working cycles can, on account of the fuel mass accumulated in the meantime, result in very high peak pressures above 150 bar.

Friction losses

Because the cold-start phase does not provide for enough time for an adequate hydrodynamic lubricating film to be built up at the appropriate bearing points of the crankshaft drive, the result is a negative effect on the mechanical engine systems. Typical measures to help improve the cold-start process include therefore preheating the intake air, the oil or the water. The latter, as well as influencing the combustion-chamber temperature, above all brings about lower engine friction and increased starter speeds.

Height dependence

A further phenomenon is encountered during operation at very hot outside temperatures or at high altitudes above 1000 m. Because the air has a lower density, its mass in the cylinder is reduced. The effect on combustion is initially not decisive. However, the exhaust-gas temperature rises as a result of the lower amount of excess air.

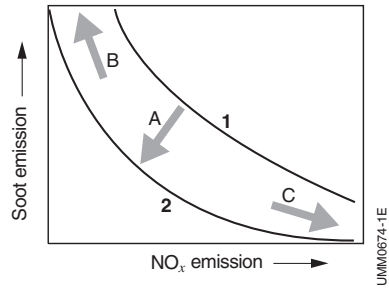
This phenomenon is also encountered with turbocharged engines. A power reduction above all during operation at altitude can therefore be a necessary application measure.

Deposits in the injection nozzles

A power drop in the order of 1 to 3 % is often observed in diesel engines after a break-in period. This is caused by the fuel-injection system. Deposits in the injection nozzles cause a slightly reduced nozzle-hole diameter, resulting in a smaller mass flow and thus a power loss. These deposits may be caused, for example, by copper, zinc or oil contaminants in the diesel fuel.

Figure 19: NO_x and soot emissions

1 Upper limit curve,
2 Lower limit curve.
A Decrease in soot and NO_x as a result of: exhaust-gas recirculation and injection-pressure increase, partial homogenization, H_2O injection.
B Increase in soot and decrease in NO_x as a result of: ignition retard, injection-pressure reduction, reduction of O_2 concentration, Miller process.
C Decrease in soot and increase in NO_x as a result of: ignition advance, injection-pressure increase, increase in O_2 concentration.



Pollutant formation and pollutant reduction

In contrast to gasoline engines, which are characterized with the introduction of the three-way catalytic converter during operation at $\lambda = 1$ by extremely low emissions, pollutant reduction inside the engine plays a much more important role in diesel engines. In addition to the emissions familiar from gasoline engines, i.e. CO_2, H_2O, NO_x, HC and CO, soot and particulate emissions must also be taken into consideration.

Pollutants

Nitrogen oxides

Measures which lower the combustion temperatures are useful in reducing the nitrogen oxides. This can be done effectively by reducing the oxygen concentration in the combustion zone. The combustion temperatures can also be reduced very easily by retarding injection or reducing the injection pressure.

Temperature-reducing measures such as exhaust-gas recirculation, Miller processes (see e.g. [2]) or partial homogenization reduce the nitrogen-oxide levels. This overcompensates for the slight increase in soot emissions that are often observed as well (Figure 19). The level of complexity and expenditure needed to reduce both these components is high. The measure currently used involves increasingly higher exhaust-gas recirculation rates for nitrogen-oxide reduction in combination with very high injection pressures (> 2000 bar) for soot reduction.

Soot and particulates

Reducing the injection pressure or the oxygen concentration typically causes an increase in soot emissions. The formation of soot is a complex process and is dependent on both fluid-dynamic and thermodynamic boundary conditions. Initially, the locally highly rich zones ($\lambda << 1$) produce considerable amounts of soot, which are then however substantially reduced by over 70 % in the subsequent combustion by oxidation processes. A high turbulence level, which encourages the soot to oxidize in the expansion phase, is crucially important. But the temperature level too is important. On account of important local interactions between the injection spray, the combustion zone, the non-combusted mixture and the recess geometry, how the combustion process is fashioned has a decisive impact on the production of emissions.

A distinction is made between soot and particulate emissions in this context. Soot consists of pure carbon compounds, whereas particulates also include fuel or lube-oil droplets, ash, rubbed-off metal, corrosion products, and sulfate compounds.

HC and CO compounds

HC and CO compounds are typically uncritical in diesel-engine emissions. The effect of hydrocarbons on particulate emissions must be taken into consideration. HC and CO increase in concentration in particular during a highly retarded combustion with incomplete burn behavior.

Mixed forms and alternative operating strategies

The classic diesel-engine operating strategy is characterized by one or more injections in the TDC range. Established gasoline-engine combustion processes are characterized by homogeneous or partially homogeneous (with charge stratification) operation. However, alternative forms of process control, which sometimes cannot be clearly assigned to the diesel-engine or the gasoline-engine process, are currently being developed.

Homogeneous-charge compression ignition in diesel operation
With the HCCI combustion process (Homogeneous Charge Compression Ignition), which is published in a wide range of variations [3], the motivation is, through the application of very advanced injection (at least 40 to 50° CA before TDC), to achieve homogenization, significant leaning and therefore a NO_x reduction. Reliable ignition nevertheless takes place on account of the high diesel-engine compression temperature. The compression ratio must be lowered to 14 to 16 in order to maintain control of the combustion process. Exhaust-gas recirculation is typically used to enable the temperature level in the cylinder to be increased at low loads. Nevertheless, obtaining this over the entire program map, particularly in the full-load range, is very demanding, since the pressure-increase gradients become extremely high and control of transient engine operation is highly complex in view of all the engine states that arise.

Compression ignition in gasoline operation
Gasoline-engine combustion processes have been further developed along similar lines to diesel-engine HCCI operation in order to achieve in the part-load range dethrottled lean-burn operation with corresponding fuel-consumption advantages over conventional, stoichiometrically operated engines. The disadvantages of lean-burn operation with regard to conversion in the catalytic converter are compensated by extremely low untreated NO_x emissions caused by leaning. Reliable ignition of the in itself hardly combustible mixture is achieved by a high compression ratio above 13. Optimally the compression ratio is variable and can be further decreased up to full load on account of the increased combustion-chamber temperatures.

Gasoline-engine stratified charge
Direct-injection, stratified-charge gasoline-engine combustion processes share much in common in terms of combustion with classic diesel-engine operation and therefore represent a mixed form of conventional gasoline-engine and diesel-engine combustion. Combustion processes of this type are becoming increasingly established thanks to their efficiency advantages by dethrottling in part-load operation.

Multifuel engines
Multifuel engines, which are characterized by compatibility with regard to the fuel used (alternatively operation e.g. with diesel fuels or kerosene, gasolines or diesel fuels and vegetable oils), no longer play a role today since the emission requirements for such engine concepts are no longer feasible.

References
[1] H. Oertel; M. Böhle; U. Dohrmann: Strömungsmechanik. 5th Ed., Vieweg+Teubner, 2008.
[2] R. van Basshuysen; F. Schäfer (Editors): Handbuch Verbrennungsmotor, 4th Ed., Vieweg+Teubner, 2007.
[3] K. Boulouchos: Strategies for Future Combustion Systems – Homogeneous or Stratified Charge? SAE 2000-01-0650.

Charge cycle and supercharging

Charge cycle

In combustion engines employing open processes and internal combustion, the charge cycle (exhaust and refill) system must serve two decisive functions:
– Replacement is employed to return the gas medium to its initial (start of cycle) condition,
– The oxygen required to burn the fuel is provided in the form of fresh air.

The charge cycle – also called gas exchange – can be assessed with the aid of the characteristic quantities defined in DIN 1940 [1]. For overall air flow (air expenditure $\lambda_a = m_g/m_t$) the entire charge transferred during the working cycle m_g is defined with reference to the theoretical maximum m_t for specific displacement. In contrast, the volumetric efficiency $\lambda_{a1} = m_z/m_t$ is based exclusively on the fresh charge m_z actually present or remaining in the cylinder. The difference between this and the total charge transfer m_g consists in the proportion of gas that flows directly into the exhaust tract in the overlap phase, making it unavailable for subsequent combustion.

The retention rate $\lambda_a = m_z/m_g$ is an index of the residual charge in the cylinder.

The scavenge efficiency $\lambda_S = m_z/(m_z + m_r)$ indicates the volume of the fresh charge m_z relative to the existing total charge, consisting of the fresh charge and the residual gas m_r. Here, the parameter m_r indicates the amount of residual gas from earlier working cycles remaining in the cylinder after the exhaust process.

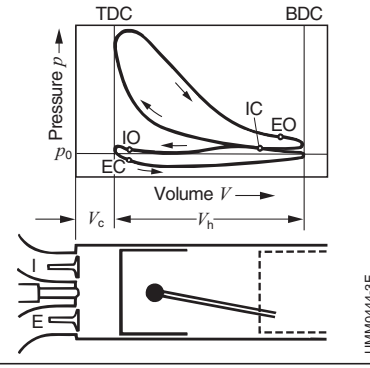
In a 2-stroke cycle, the gas is exchanged with every rotation of the crankshaft at the end of the expansion in the area around bottom dead center. In a 4-stroke cycle, separate intake and exhaust strokes provide a supplementary gas-exchange cycle.

4-stroke cycle

Valve timing – and thus gas exchange – are regulated by a control shaft (camshaft) rotating at half the frequency of the crankshaft by which it is driven. The camshaft opens the gas-exchange valves by depressing them against the valve springs to discharge the exhaust gas and to draw in the fresh gas (exhaust and intake valves respectively) (Figure 1). Just before piston bottom dead center (BDC), the exhaust valve opens and approx. 50 % of the combustion gases leave the combustion chamber under a supercritical pressure ratio during this predischarge phase. As it moves upward during the exhaust stroke, the piston sweeps nearly all of the combustion gases from the combustion chamber.

Shortly ahead of piston top dead center (TDC) and before the exhaust valve has closed, the intake valve opens. This crankshaft top dead center position is called the gas-exchange TDC (GTDC) or overlap TDC (OTDC) because the intake and exhaust processes overlap at this point) in order to distinguish it from the ignition TDC (ITDC). Shortly after gas-

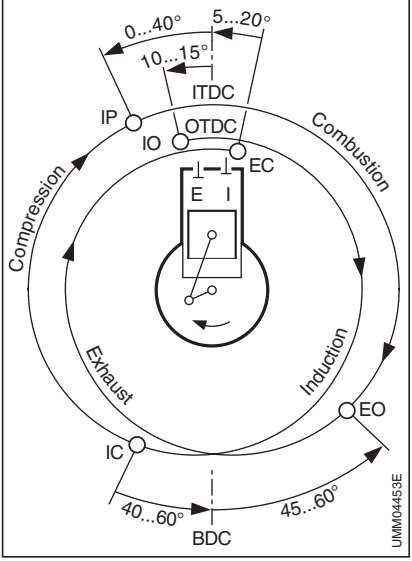
Figure 1: Representation of the 4-stroke gas-exchange process in the pV diagram
I Intake, E Exhaust,
BDC Bottom dead center, TDC Top dead center,
IC Intake closes, IO Intake opens,
EC Exhaust closes, EO Exhaust opens,
V_c Compression volume, V_h Displacement.
The arrows indicate the direction of the curves.



exchange TDC, the exhaust valve closes and, with the intake valve still open, the piston draws in fresh air on its downward stroke. This stroke of the gas-exchange process, the induction cycle, is completed shortly after bottom dead center (BDC) is reached. The subsequent two strokes in the 4-stroke process (Figure 2) are compression and combustion (expansion).

On throttle-controlled gasoline engines during the valve overlap period, exhaust gases flow directly from the combustion chamber into the intake passage, or from the exhaust passage back into the combustion chamber and from there into the intake passage. This internal exhaust-gas recirculation increases the combustion-chamber-charge temperature and the proportion of inert gas in the cylinder. A consequence of this is non-optimum power utilization in the upper load range. With fixed valve timing it is important to maintain a compromise, which is dependent above all in gasoline-engine applications on matching the engine operating strategy with the turbocharging unit.

Figure 2: 4-stroke gas-exchange process
E Exhaust, I Intake,
EO Exhaust opens, EC Exhaust closes,
IO Intake opens, IC Intake closes,
TDC Top dead center,
OTDC Overlap TDC, ITDC Ignition TDC,
BDC Bottom dead center,
IP Ignition point.



Early exhaust-valve timing allows sufficient time for the cylinder charge to escape, and thus guarantees low residual-gas compression as the piston sweeps through its upward stroke, although at the price of a reduction in the work index for the combustion gases.

The “intake valve closes” position (IC timing) exercises a decisive effect on the relationship between volumetric efficiency and engine speed. The maximum volumetric efficiency is achieved at low engine speeds when the intake valve closes early, and at high engine speeds when it closes late.

Obviously, fixed valve timing will always represent a compromise between two different design objectives: maximum brake mean effective pressure – and thus torque – at the most desirable points on the curve, and the highest possible peak output. The higher the engine speed at which maximum power occurs, and the wider the range of engine operating speeds, the less satisfactory will be the ultimate compromise. This tendency cannot be neutralized either by multi-valve cylinder heads with a larger intake-flow cross-sectional area.

At the same time, the demands for minimum exhaust emissions and maximum fuel economy mean that low idle speeds and high low-end torque (despite and along with high specific outputs for reasons of power-unit weight) are becoming increasingly important. These imperatives lead to the application of variable valve timing.

The advantages of the 4-stroke process are very good volumetric efficiency over the entire engine-speed range, low sensitivity to pressure losses in the exhaust system, and relatively good control of the speed-sensitive air expenditure through selection of appropriate valve timing and intake-system designs.

The disadvantages of the 4-stroke process are the high complexity of valve timing and the reduced power density of non-turbocharged 4-stroke engines because only every second crankshaft revolution is used for power output by combustion.

2-stroke cycle

To maintain gas exchange without an additional crankshaft rotation, the gases are exchanged in the two-stroke process at the end of expansion and at the beginning of the compression stroke. The intake and exhaust timing are usually controlled by the piston as it sweeps past the intake and exhaust ports in the cylinder housing near BDC (Figure 4). This configuration, however, requires symmetrical control times and involves the problem of directly flushing fresh mixture into the exhaust area (short-circuit scavenging). In addition, 15 to 25 % of the piston stroke cannot produce work because only charge volume V_f and not displacement volume V_h can be exploited for power generation (Figure 3).

As the two-stroke process lacks separate intake and exhaust strokes, the cylinder must be filled and scavenged using positive pressure, necessitating the use of scavenging pumps. In an especially simple and very frequently-used design, the bottom surface of the piston works in conjunction with a crankcase featuring a minimal dead volume to form a scavenging pump. Figure 4 shows a 2-stroke engine with crankcase scavenging and precompression along with the associated control processes. The processes which take place on the scavenging pump side are shown in the inner circle, while those occurring on the cylinder side are shown in the outer circle. The location of the intake and exhaust ports with the aim of maximum volumetric efficiency is also crucially dependent on the piston position. A hump-shaped increase in piston height can reduce direct throughflow from the intake area to the exhaust area.

The advantages of the 2-stroke process are high power output referred to engine weight and engine volume and a more uniform torque generation (one power stroke per revolution).

The disadvantages of the 2-stroke process are higher fuel consumption, lower mean pressures (due to poorer cylinder charge), higher thermal loads (due to the lack of a gas-exchange stroke), demanding mixture control, and higher HC emissions due to problematic cylinder scavenging. Two-stroke processes are no longer important due to the introduction of strict emission limits for most mobile applications.

Figure 3: Representation of the 2-stroke gas-exchange process in the pV diagram

p_0 Charge pressure, V_S Scavenging volume, V_c Compression volume, V_h Displacement, V_f Charge volume.

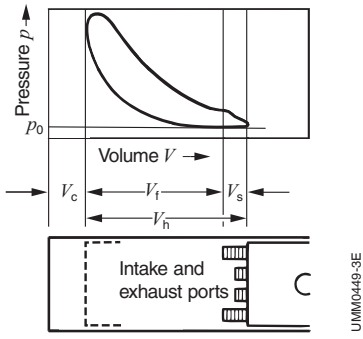
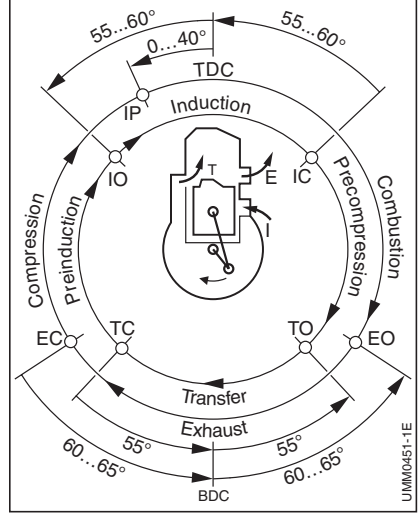


Figure 4: 2-stroke gas-exchange process with crankcase precompression

E Exhaust, I Intake, EO Exhaust opens, EC Exhaust closes, IO Intake opens, IC Intake closes, T Transfer passage, TO Transfer passage opens, TC Transfer passage closes, TDC Top dead center, BDC Bottom dead center, IP Ignition point.



Variable valve timing

Variable valve timing in a number of design variations is gaining acceptance with different objectives. The motivations behind the use of variable valve timing are as follows: increased power and torque, dethrottling, control of the residual-exhaust-gas content, cylinder shutoff, charge movement, cold-starting and cold-running performance, control-engineering optimization (e.g. charge mass, residual-exhaust-gas content), influencing the rotational speed of the exhaust-gas turbocharger, and exhaust-gas temperature-management measures.

Today, different systems are used to provide variable valve timing. There are different motivations behind variable valve timing, depending on the combustion process and engine concept (homogeneous- or stratified-charge operation, diesel).

Camshaft adjustment

The concept of adjusting the intake and exhaust camshafts is increasingly becoming an accepted measure, offering a valuable degree of variability in the entire program map. This concept provides the option of adjusting the phase of the camshafts with respect to the crankshaft and thereby shifting the valve timing while maintaining the same form of opening. Electric or electrohydraulic activation of the camshaft adjuster is the state-of-the-art technology used.

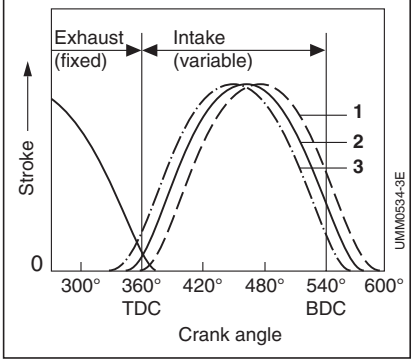
In homogeneously operated gasoline engines the intake camshaft is subject to retarded adjustment for example in the low speed and load range while at the same time the exhaust valve is subject to very advanced closing (Figure 5). This minimal valve overlap delivers a minimal throughflow of fresh-air mixture into the exhaust area. At the same time the highly retarded opening of the intake valve also causes dethrottling by the associated retarded closing of the intake valve, because the cylinder charge is again discharged by the intake valve after bottom dead center. A rapid torque increase can be achieved for a desired load point by advancing the intake camshaft.

Exhaust-camshaft adjustment also provides for further degrees of freedom. Optimal adjustment of the intake and exhaust camshafts is dependent on numerous factors. The operating mode is crucial as well as the position in the engine map. For example, a significant factor is whether exhaust-gas turbocharging is used or whether lean-burn operation (diesel engine, gasoline-engine stratified-charge process, etc.) is applied. One example of the beneficial combination of camshaft adjustment with an exhaust-gas turbocharger is the map-dependent maximum valve overlap (very retarded closing of the exhaust valve and very advanced closing of the intake valve), at which a large proportion of mixture flows directly from the intake to the exhaust area. This can give rise to the beneficial effect of an increase in air-mass flow by the exhaust-gas turbocharger together with an advantageous increase in turbocharger speed. Rotating the intake and exhaust camshafts offers optimization opportunities for a large number of questions and is therefore being increasingly used at least in modern gasoline engines.

Systems with cam-shape control mechanism

As well as continuous phase adjustment, simpler systems are also available which alternate digitally between individual cam shapes. Typically, such systems adjust a very large cam and valve lift for ranges

Figure 5: Intake-camshaft rotation
1 Retard, 2 Standard, 3 Advance.



close to full load and a small cam and valve lift for dethrottling for part-load ranges. This is achieved for the most part by integrating a control mechanism on a rocker arm or bucket tappet, which thereby activates one or more (typically two) available cams on the camshaft.

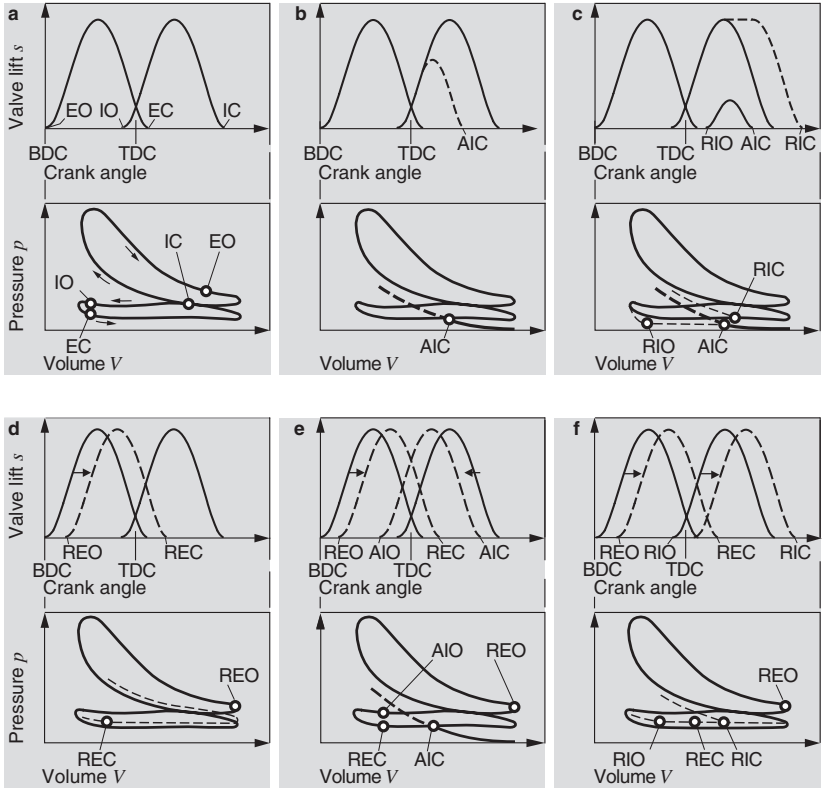
Systems that can alternate between two camshaft rotation positions are also used. In contrast to fully variable systems, these systems omit an exact positioning command to the camshaft phase adjuster.

Other variable systems

Increasingly, fully variable systems for use in gasoline engines are being developed and are even already appearing in production vehicles. These systems focus on optimizing the charge-cycle phase. There are different successful approaches to suit the engine configuration and operating point, each approach optimizing charge-cycle losses, the residual-exhaust-gas content or engine power (Figure 6).

Figure 6: Influence of valve timing on the charge cycle

a) Conventional valve timing, b) Advanced "Intake closes",
c) Valve-lift reduction and retarded "Intake closes", d) Retarded "Exhaust closes",
e) Retarded "Exhaust closes" and advanced "Intake closes",
f) Retarded "Exhaust closes" and retarded "Intake closes".
BDC Bottom dead center, TDC Top dead center, IO Intake opens,
EO Exhaust opens, IC Intake closes, EC Exhaust closes,
A Advance, R Retard.



Mechanical, electromechanical, electrohydraulic or electropneumatic adjustment concepts can be used to provide a fully variable valve gear.

Mechanical systems
Mechanical, fully variable valve-gear systems typically consist of a combination of an adjustment mechanism, which permits variation of the valve lift, and a camshaft phase adjuster. The key function is the provision of variable cam lift.

A suitable design solution is provided in this case by a conventional camshaft, which does not however act directly on the valve via a bucket tappet or a rocker arm. The cam lift is adjusted by an intermediate lever whose fulcrum can be varied by means of an eccentric shaft (Figure 7 and Figure 8). Electric DC motors are typically used to drive such arrangements. BMW Valvetronic is an example of a fully variable mechanical valve gear [2]. Further advantages of a fully variable valve gear are the optimization of uniform charge distribution and the potential to increase charge movement by having the two intake valves opening at different times in the lower load range thanks to slightly differently ground eccentricities.

Figure 7: Valve-lift curves of a mechanical fully variable valve gear

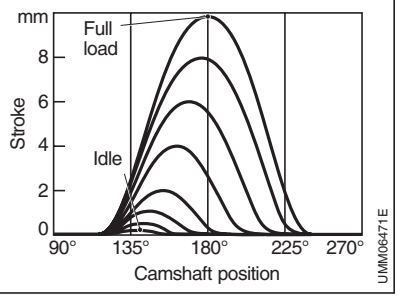
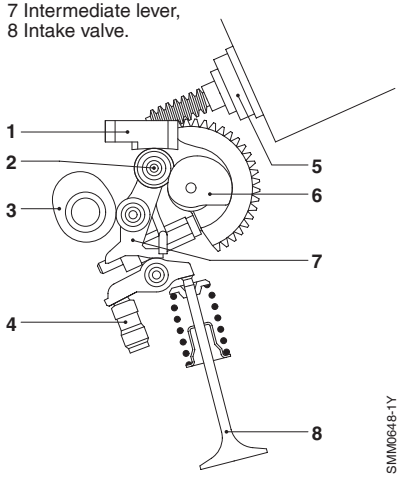
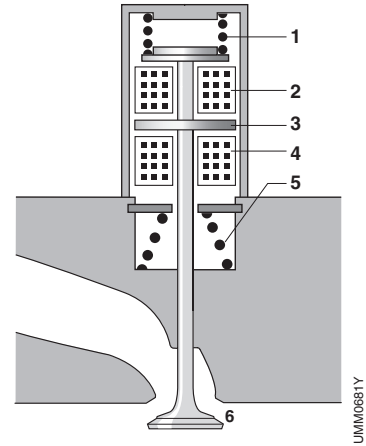


Figure 8: Basic principle of a mechanical fully variable valve gear (BMW Valvetronic)
1 Guide block,
2 Center of rotation,
3 Camshaft,
4 Hydraulic valve-clearance compensation,
5 Servo unit with spiral-toothed gear,
6 Eccentric shaft,
7 Intermediate lever,
8 Intake valve.



Electromechanical systems
Electromechanical systems (electromechanical valve gear) are still in the development stage. In an electromechanical valve gear, electrically actuated magnets are used as actuators for valve timing (Figure 9, see also [3]). The high power

Figure 9: Electromechanical systems
1 Compression spring,
2 Closing magnet,
3 Armature,
4 Opening magnet,
5 Compression spring,
6 Intake valve.



demand of the actuators must be noted in particular. The system, comprising valve, magnet and coil, is made to resonate in order to minimize this demand. Because of the high power demand and the complexity involved, electromagnetic valve gears have not yet made it to series production.

Electrohydraulic systems

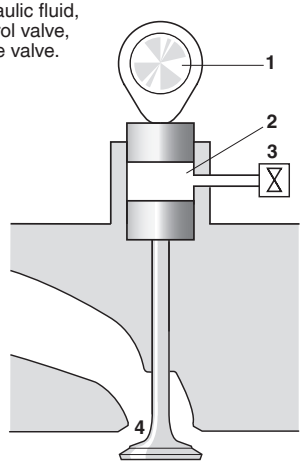
Electrohydraulic systems (electrohydraulic valve timing) are an alternative to the fully variable mechanical valve gear. Different principles are applicable here.

The principle of "lost motion" is an effective approach. A motion is stipulated by a hydraulic intermediate element via a camshaft (Figure 10). An electrically actuated hydraulic valve provides the opportunity of not transferring completely the motion stipulated by the cam. The cam shape therefore stipulates an envelope (Figure 11).

An alternative is presented by a system which directly actuates the gas-exchange valves via a hydraulic pressure accumulator and electronically controlled hydraulic valves. Systems of this type are still in the development stage and are not yet in series production.

Figure 10: Working principle of electrohydraulic systems

1 Camshaft,
2 Hydraulic fluid,
3 Control valve,
4 Intake valve.



Depending on the point and duration of activation, a proportion of the hydraulic fluid escapes through the hydraulic valve. Hydraulic power loss is therefore a drawback of this system.

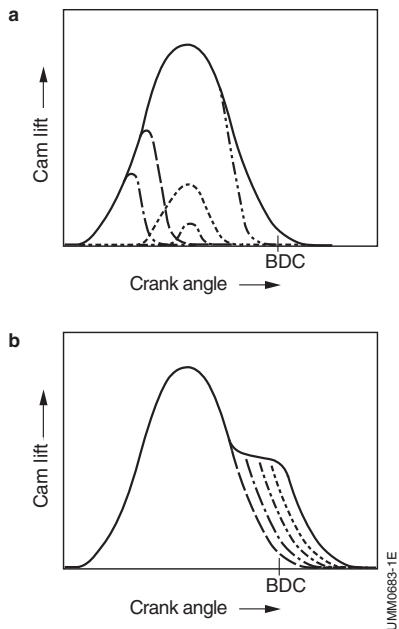
Electrohydraulic systems have been deployed since 2004 in Caterpillar commercial-vehicle systems and since 2010 in Fiat's MultiAir system.

Electropneumatic systems

Electropneumatic systems are still in the development stage. The use of such systems in large-scale production is not anticipated at present. As well as exhibiting complex control characteristics, these systems above all demonstrate unfavorable pneumatic drive power. Current systems are used for research purposes. The power input of the compressed-air supply must be noted and necessarily included in an efficiency balance.

Figure 11: Cam lifts of electrohydraulic systems for different applications

a) Advanced closing of intake valve,
b) Variable retarded closing of intake valve.



Supercharging processes

The power of an engine is proportional to its air-mass flow. As this air throughput, in turn, is linearly dependent on air density, the power of an engine, given a specific displacement and engine speed, can be increased by compressing the air before it enters the cylinders, i.e. by supercharging. The supercharging ratio indicates the density rise as compared to a naturally aspirated engine.

Thermodynamically speaking, the best results would be delivered by isothermal compression, but this cannot be technically achieved. Reversibly adiabatic compression in the supercharger unit serves as the ideal comparison process for the actual sequences; in practical terms, increased density is accompanied by losses.

The extent of the supercharging ratio is limited in gasoline engines by the occurrence of knocking combustion and in diesel engines by the maximum permissible peak pressure in the cylinder. To offset this limitation, supercharged engines usually have lower compression ratios than their naturally aspirated counterparts.

Dynamic supercharging

The gas-exchange processes are not only influenced by the valve timing, but also by the geometry of the intake and exhaust lines. Induced by the induction work of the piston, as the intake valve opens it initiates a suction wave in the intake manifold that reflects off the open end of the intake manifold and returns as a pressure wave to the intake valve. These pressure waves can be used to increase the air-mass intake (Figure 12). In addition to the geometry in the intake manifold, the impact of this supercharging effect based on gas dynamics also depends on the engine speed (see Ram-tube supercharging, Tuned-intake-tube charging, Variable-geometry intake-manifold systems).

Mechanical supercharging

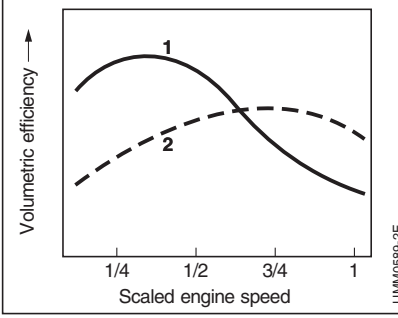
In mechanical supercharging, the supercharger is driven directly by the internal-combustion engine (see Turbochargers and superchargers). The charger and the internal-combustion engine are mechanically coupled to each other. Types of mechanical supercharger unit are positive-displacement superchargers (compressors) of various designs (e.g. Roots superchargers, spiral-type superchargers) and hydrokinetic flow compressors (e.g. radial compressors).

The crankshaft and charger shaft feature a fixed transmission ratio in system designs used up to now. Mechanical or electromagnetic clutches can be used to control supercharger activation. The boost pressure is generally adjusted by a bypass device with control flap (wastegate).

Advantages of mechanical supercharging
- The supercharger is installed on the cold side of the engine.
- The engine's exhaust system is not influenced by supercharger components.
- The supercharger responds almost immediately to load changes.

Disadvantages of mechanical supercharging
- The drive power must be branched from the effective power of the engine, which results in increased fuel consumption.

Figure 12: Increase of volumetric efficiency through dynamic supercharging
The engine speed is scaled to the nominal speed.
1 System with dynamic supercharging,
2 System with standard intake manifold.



– Low-noise performance can only be achieved by additional measures.
– Comparatively high structural volume and weight.
– The supercharger unit has to be positioned at the engine's belt level.

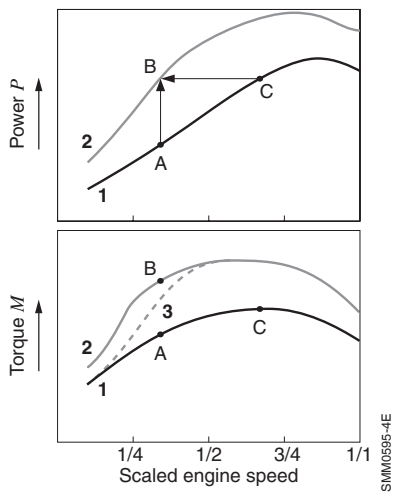
Exhaust-gas turbocharging

In exhaust-gas turbocharging, the power for driving the charger is extracted from the engine's exhaust gas, i.e. some of the energy contained in the exhaust gas is converted into mechanical energy by an exhaust-gas turbine. In this way, the process exploits some of the enthalpy that remains unused (owing to crankshaft-assembly expansion limits) by naturally aspirated engines. Exhaust backpressure however increases in the process. Hydrokinetic flow compressors are exclusively used to compress the induction air (see Turbochargers and superchargers).

Exhaust-gas turbochargers are usually designed to generate a high boost pressure even at low engine speeds. In other words, the turbocharger turbine is generally designed for an engine mid-speed range. However, without further measures, boost pressure at the high end of the engine speed range could increase to levels that would place excessive load on the engine. The turbine is therefore equipped with a bypass valve which, from a specific operating point, diverts some of the exhaust-gas mass flow past the turbine. However, the energy of this exhaust gas remains unutilized. Much more satisfactory results can be achieved with a compromise between high boost pressure in the low engine-speed range and avoidance of engine overload at the high end of the engine-speed range by employing a turbocharger with variable turbine geometry (VTG). The adjusting blades used for this purpose adapt the flow cross-section and the angle of impact to the turbine (and thus the exhaust-gas pressure applied to the turbine) by varying the position of the guide blades (see Turbochargers and superchargers).

Figure 13: Comparison of power and torque curves between naturally aspirated engine and turbocharged engine

The engine speed is scaled to the nominal speed.
1 Naturally aspirated engine in stationary operation,
2 Turbocharged engine in stationary operation,
3 Torque build-up of a turbocharged engine in transient operation.
A→B Higher torque and higher power at identical speed for turbocharged engine.
C→B Identical power for turbocharged engine at lower speed.



Advantages of exhaust-gas turbocharging
– Considerable increase in power output per liter.
– Improvement in fuel-consumption figures relative to naturally aspirated engines with the same output power.
– Improvement in exhaust-gas emission rates.
– Comparatively small structural volume.
– Can be used with low-pressure-side exhaust-gas recirculation.

Disadvantages of exhaust-gas turbocharging
– Installation of the turbine side of the turbocharger in the hot exhaust-gas system requiring materials resistant to high temperatures.
– Increased thermal inertia in the exhaust-gas system.
– Without further measures, comparatively low starting torque from low-displacement engines.

Special forms of supercharging

In electrically assisted exhaust-gas turbocharging systems, an additional electric motor serves to maintain the turbocharger at a high speed in the absence of exhaust-gas mass flow. The produces benefits above all in transient vehicle operation and at low engine speeds, but the complexity and the electric power consumption are very high. The drawbacks of the exhaust-gas turbocharger due to inertia can be reduced with electrically assisted exhaust-gas turbocharging.

A special but only theoretical form of supercharging is pressure-wave supercharging (Comprex); this however failed with regard to production maturity. The working principle is based on the reflection properties of pressure waves which propagate in a rotating cell rotor (see Turbochargers and superchargers). The chief advantage is the transient response, which facilitates a very rapid torque build-up. However, the costs involved with pressure-wave superchargers are high and the required drive necessitates additional packaging measures.

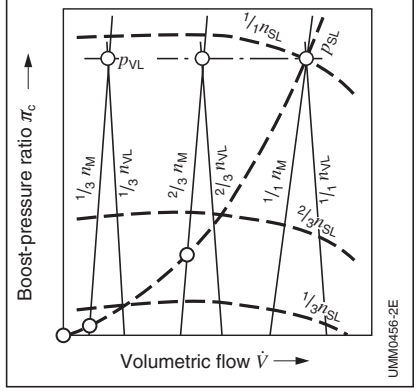
Volumetric-flow map

A clear illustration of the relationship between the engine and the supercharger is provided by the diagram for pressure vs. volumetric flow (Figure 14), in which the pressure ratio π_c of the supercharger is plotted against volumetric flow $\dot{V}$.

The curves for unthrottled 4-stroke engines (diesel) are particularly descriptive because they contain sloped straight lines (engine mass-flow characteristics) which represent increasing air-throughput values as the pressure ratio $\pi_c = p_2/p_1$ (where p_1 denotes the ambient pressure and p_2 the charge-air pressure) increases at constant engine speed.

Figure 14: Pressure vs. volumetric-flow map of mechanically driven positive-displacement supercharger and hydrokinetic flow compressor

n Speed, p_L Boost pressure.
Subscripts:
VL Positive-displacement supercharger,
SL Hydrokinetic flow compressor,
M Engine.



The diagram shows pressure ratios which result at corresponding constant supercharger speeds for a positive-displacement supercharger and a hydrokinetic compressor.

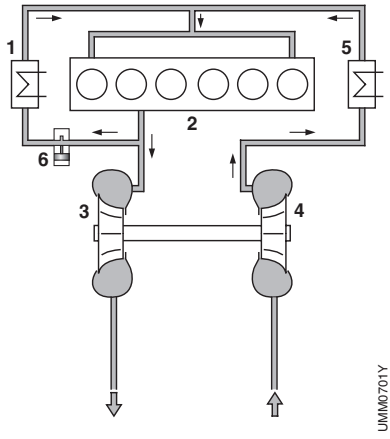
Only superchargers whose delivery rates vary linearly with their rotational speeds are suitable for vehicle engines. These are positive-displacement superchargers of piston or rotating-vane design or Roots blowers. Mechanically driven hydrokinetic flow compressors are not suitable.

Exhaust-gas recirculation

Operating principle
With exhaust-gas recirculation (EGR) hot exhaust gas is removed, cooled in an exhaust-gas cooler, mixed with fresh air, and returned to the combustion chamber. The exhaust-gas recirculation rate is adjusted by a valve. It is used in diesel engines and also increasingly in gasoline engines.

Exhaust-gas recirculation in diesel engines
In diesel engines the untreated emissions of NO_x are reduced significantly by exhaust-gas recirculation. Essentially this is done by lowering the peak temperature in the cylinder, which is obtained by the fact that the on the one hand the recirculated exhaust gas reduces the combustion velocity and on the other hand the exhaust gas on account of its now higher heat capacity – above all of the CO2 – for the same supplied energy does not become as hot as exhaust gas without exhaust-gas recirculation.

Figure 15: High-pressure exhaust-gas recirculation
The arrows denote the flow direction.
1 EGR cooler for high-pressure EGR,
2 Engine,
3 Exhaust-gas-turbocharger turbine,
4 Exhaust-gas turbocharger compressor,
5 Intercooler,
6 EGR valve.



Exhaust-gas recirculation in gasoline engines
In gasoline engines the idea is not to reduce the nitrogen oxides. The aim is primarily to reduce the exhaust-gas temperature to protect components (catalytic converter and turbocharger). This is very effective particularly in engine full-load operation because the otherwise usual process of enriching the air/fuel mixture to reduce the exhaust-gas temperature can be partially or completely dispensed with. This results in a dramatic improvement in fuel consumption.

An additional consideration in gasoline engines the possibility of diluting the charge by means of exhaust-gas recirculation without compromising the function of the three-way catalytic converter. Engine dethrottling can be achieved by increasing the EGR rate particularly in engine part-load operation, which also results in improved fuel efficiency.

Exhaust-gas recirculation variants
Essentially, one of two variants – high-pressure and low-pressure recirculation – is used, depending on whether the exhaust-gas cooler is located upstream or downstream of the turbocharger. Whereas high-pressure recirculation is found primarily in commercial vehicles, both variants are used in passenger cars.

High-pressure EGR
With high-pressure recirculation (Figure 15), exhaust gas is recirculated on the high-pressure side, i.e. upstream of the turbocharger turbine. It has the advantage that the pressure difference required to transport the exhaust gas from the exhaust side to the intake side is higher and can even be altered by means of variabilities at the turbocharger (wastegate or variable turbine geometry). High EGR rates can basically be achieved because the EGR section is short and can therefore be filled and emptied quickly. Thanks to this dynamic advantage, the EGR system can respond relatively quickly to changing operating conditions.

Because the recirculated exhaust gas is not mixed with fresh air until downstream of the intercooler, the intercooler and the turbocharger compressor are not

permeated and thus not fouled by the exhaust gas. The performance of these components therefore remains constant over their full service lives. By contrast, significantly greater demands are placed on the EGR cooler on the high-pressure side. This can be put down to the higher pressures and temperatures during operation and to the already mentioned component sooting.

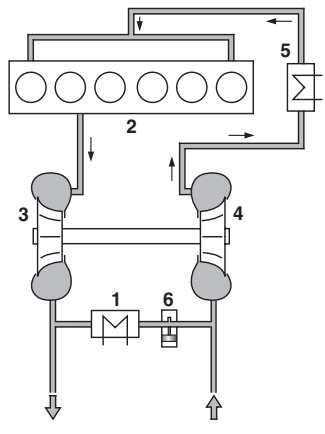
Low-pressure EGR

This form of exhaust-gas recirculation routes the exhaust gas only downstream of the turbine through the EGR cooler and mixes it with the inducted fresh air upstream of the turbocharger compressor (Figure 16). The exhaust gas arrives at the EGR cooler only after it has passed through the diesel particulate filter (DPF, not shown in the illustration) and the soot has been removed. Thus the EGR cooler is not fouled and performs consistently over its full service life. Because of the lower temperature in this area of the exhaust system the temperature stresses in the EGR cooler are also much lower than is the case for high-pressure recirculation. The lack of fouling together with a lower inlet temperature provide for a lower cooler outlet temperature than with high-pressure recirculation. The reduction in NO_x emissions that can thereby be achieved can be sufficient in many cases to eliminate the need for further exhaust-gas treatment, such as the SCR system for example.

One drawback of low-pressure recirculation is the reduced dynamic response. The system is relatively far removed from the engine and, due to the long distances involved, cannot adapt the exhaust-gas recirculation rate quickly. It is necessary to set the EGR rates slightly lower than for high-pressure recirculation so as to prevent the engine from being "flooded" with exhaust gas.

Figure 16: Low-pressure exhaust-gas recirculation

The arrows denote the flow direction.
1 EGR cooler for low-pressure EGR,
2 Engine,
3 Exhaust-gas-turbocharger turbine,
4 Exhaust-gas turbocharger compressor,
5 Intercooler,
6 EGR valve.



References

[1] DIN 1940: Reciprocating internal combustion engines; terms, formulae, units.

[2] R. Flierl; R. Hofmann; C. Landerl; T. Melcher; H. Steyer: Der neue BMW-Vierzylindermotor mit Valvetronic. MTZ 62 (2001), Volume 6.

[3] P. Langen; R. Cosfeld; A. Grudno; K. Reif: Der Elektromechanische Ventiltrieb als Basis zukünftiger Ottomotorkonzepte. 21st International Vienna Motor Symposium, Progress Reports VDI, Series 12, No. 420, Volume 2, 2000.

[4] R. van Basshuysen; F. Schäfer (Editors): Handbuch Verbrennungsmotor, 5th Edition, Vieweg-Verlag, 2009.

Reciprocating-piston engine

Components

In addition to the crankshaft drive, which makes a vital contribution by generating torque, the significance of the cylinder head is crucial to the efficiency of an internal-combustion engine (Table 1).

Crankshaft drive and crankcase
The crankshaft drive comprises the following components: piston, piston rings, connecting rod, and crankshaft. What all the crankshaft-drive components have in common is a translational and rotational motion. The tribological design of the crankshaft drive is already hugely significant in view of its robustness and reduced engine friction.

Piston
Because of the mechanical and thermal loads to which it is subjected, the piston is a highly complex component in an internal-combustion engine. The recess and piston-surface geometries have a crucial impact on mixture formation and combustion. In addition, the primary function of the piston is to transmit force to the connecting rod. Complex mechanical stresses occur in the process with simultaneously very high local temperatures up to and in excess of 300 °C.

The role of the piston recess is even more important in diesel engines. With the high compression ratio and a typically flat cylinder head, the entire combustion chamber must be accommodated in the piston recess. Various piston-recess designs are successfully used. Figure 1 shows examples of the design variations used in diesel engines.

The piston recess decisively defines the combustion-chamber shape and at the same time influences the strength of the piston crown, a factor which is vital to the function of force input into the piston pin. Especially when it comes to aluminum pistons, which are currently the preferred choice mainly for weight reasons, the requirements of high heat resistance, low weight and cylinder peak pressure call for technically sophisticated solutions.

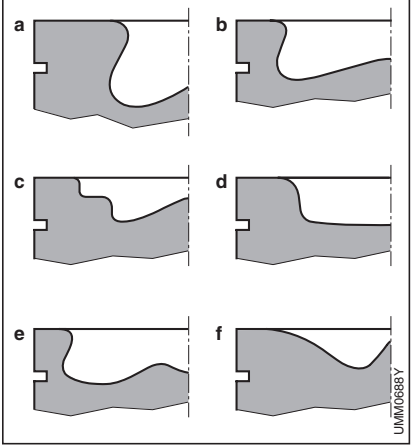
Brass bushings are used to accommodate the piston pin in diesel engines subject to high loads.

Diesel engines can also utilize steel and even gray-cast-iron pistons as well as aluminum designs. Steel pistons are used in applications with peak pressures well in excess of 200 bar. An important distinguishing feature of steel is its lower heat conduction.

Modern turbocharged engines with high cylinder peak pressures and high power outputs also require a piston spray injector with an additional piston cooling duct for cooling purposes (Figure 2). Here engine oil is sprayed through a vertically arranged piston spray injector into the piston cooling duct. The inlet and outlets for the piston cooling duct are located on the underside of the piston.

Piston rings form the sealing element between the combustion chamber and the crankcase (Figure 3). As well as providing the seal, the piston rings also perform the important tasks of conducting heat to the

Figure 1: Piston shapes for diesel engines
a) Deep omega recess,
b) Flat omega recess,
c) Stepped recess,
d) Dish recess,
e) Eccentric recess design,
f) W recess.



cylinder wall and controlling the entry of oil into the combustion chamber.

Three rings are typically used, the top ring acting as a compression ring and the bottom ring as an oil control ring. The middle ring frequently performs both functions. Because particulate filters become clogged with ash, which is also a product of burnt oil in the combustion chamber, oil consumption and therefore the coordination of the piston-ring set and honing are vitally important.

Table 1: Components of a 4-stroke reciprocating-piston engine

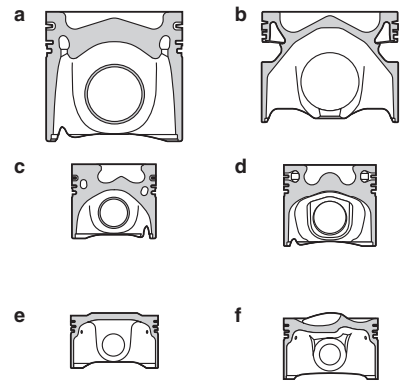
Assembly	Component	Purpose and function	Stress
Crankcase	Crankcase	– Housing for crankshaft drive – Accommodation of cylinder head and counterpart for accommodation of cylinder-head bolts – Accommodation of crankshaft bearings – Accommodation of accessories	Deformation – flow of force from cylinder-head bolts to crankshaft, bending moments, vibrations, oscillations and natural frequencies
	Bearing surface	Different approaches to structural solution of bearing surface: – Accommodation of liners with seal (wet liners) – Accommodation of cast liners (dry liners) – Piston bearing surface directly integrated in crankcase	Deformation (out-of-round), tribological counterpart to piston
Crankshaft drive	Piston	– Conversion of energy in combustion chamber (gas pressure) into a translational motion – Dissipation of heat from combustion chamber to cooling medium – Support of the normal forces occurring as a result of the connecting-rod inclination	Complex mechanical and local thermal stress above 350 °C in the recess area
	Piston rings	– Sealing of combustion chamber in relation to crankcase – Avoidance of oil entry into cylinder – Dissipation of heat to cylinder wall	High mechanical bending stress and high thermal stress, complex tribology
	Connecting rod and pin	– Absorption of the forces acting on the piston and transmission to the crankshaft for torque build-up	High mechanical compressive, tensile and bending stresses, complex tribological processes
	Crankshaft	– Absorption of the forces acting on the connecting rod and conversion of oscillatory to rotational motion. Thus build-up of a torque for transmission to the drivetrain	High mechanical compressive, tensile and bending stresses (buckling), complex tribological processes
Cylinder head	Cylinder head	– Control of charge cycle – Cooling of areas close to the combustion chamber – Accommodation of injectors (in diesel engine and gasoline engine with direct injection) and of spark plug (in gasoline engine)	High stress due to thermomechanical fatigue behavior and cold-hot expansion

Table 1: Components of a 4-stroke reciprocating-piston engine (continued)

Assembly	Component	Purpose and function	Stress
Cylinder head	Camshaft (typically in cylinder head)	– Conversion of a rotational motion into a translational motion to open and close the valves	Hertzian stress
	Valve control	– Conversion of a rotational motion into a translational motion to open and close the valves	High valve-closing speeds require large valve spring forces
	Valves	– Opening of passages to intake and exhaust ports – Sealing of combustion chamber during compression phase	Valve-seat wear due to high valve-closing speeds, thermal load at exhaust valve
	Ports/ passages	– Supply and discharge of cylinder charge – Provision of the desired cylinder-charge movement	Local thermal load
	Base plate	– Accommodation of combustion-chamber pressure and sealing	Local thermal load and bending
	Geartrain	– Driving of camshaft by belt, chain or gear wheels	Rotational irregularities with alternating contact flanks
Other components	Oil pump	– Provision of the required oil delivery rate and provisions of the required oil pressure	Torque particularly with viscous oil
	Oil filter	– Filtering of contaminants from engine oil	–
	Water pump	– Provision of the required cooling-medium delivery rate – Provision of the required cooling-medium mass flow	Cavitation

Figure 2: Piston shapes in various engine designs

a) Commercial-vehicle aluminum diesel-engine piston with ring carrier and cooling channel,
b) Commercial-vehicle forged-steel piston,
c) Passenger-car aluminum diesel-engine piston with ring carrier and cooling channel,
d) Passenger-car aluminum diesel-engine piston with cooled ring carrier,
e) Passenger-car aluminum piston for MPI gasoline engine (Multi-Point Injection),
f) Passenger-car aluminum piston for gasoline engine with GDI (gasoline direct injection).



Connecting rod

The connecting rod (Figure 4) links the piston with the crankshaft. It transmits the piston's gas and inertial forces introduced into the piston pin to the crank pin. It is therefore subjected to tensile, compressive and bending stresses. It also carries the bearings for the crankshaft and usually for the piston pin. It must therefore be particularly stiff in its design, especially in the connecting-rod eyes, which accommodate the bearings. For this reason, it is usually designed with a double-T-shaped shank profile and drop-forged with tempering steels. Sintered materials can also be successfully used. Annealed cast iron can also sometimes be used in low-load applications (gasoline engines). Rarely encountered is the combination of a standard connecting rod and a forked connecting rod in V-engines which then permits an exactly opposed cylinder arrangement.

The length of the connecting rod is determined by the piston stroke and the counterweight radius (the geometric path of the outer contour is called the "connecting-rod violin"). In commercial-vehicle engines with a relatively large stroke/bore ratio of more than 1.1 and large bearing diameters an angled split of the big connecting-rod eye is customary. This enables the piston to be removed without the crankshaft having to be removed also. However, this angled split generates considerable shear forces even when the connecting rod is positioned vertically at top dead center. These forces are usually absorbed by toothing, groove and spring or by cracked surfaces. Cracking of the big connecting-rod eye is achieved by a "fracture separation" using extremely high (hydraulic) expansion forces and a targeted weakening of the cross-section at the big connecting-rod eye. The big-end bearing cap and the connecting rod can only be installed as a single unit.

The bolts are designed as through-bolts with or without nuts. Adapter sleeves or bolt adapter collars ensure that the bearing shells are aligned when seated.

Lateral guidance of the connecting rod can be performed in the crankshaft at the big connecting-rod eye or in the piston at the small connecting-rod eye.

The small connecting-rod eye must absorb considerable forces and is therefore often form-bored in order to avoid misalignment. For diesel engines with high peak pressures in particular and in accordance with high gas forces, a special

Figure 3: Piston-ring shapes and configurations

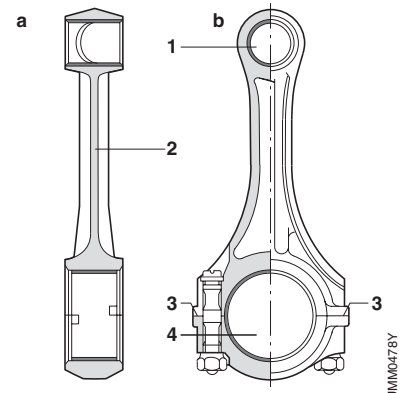
Diesel engine:
1 Keystone ring, crowned,
2 Taper-face compression ring with inside bevel,
3 Stepped compression ring,
4 Double-beveled ventilated oil control ring with spiral-type expander.

Gasoline engine:
5 Plain compression ring, crowned,
6 Taper-face compression ring,
7 Stepped ring,
8 Double-beveled ring,
9 Multipart steel oil ring.



Figure 4: Connecting rod for a passenger-car engine

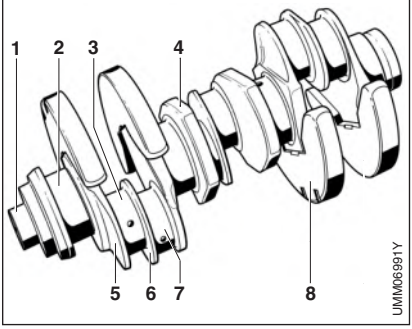
a) Sectional view, b) Side view.
1 Small connecting-rod eye,
2 T-shaped shank profile,
3 Straight split,
4 Big connecting-rod eye.



shape is chosen for the small connecting-rod eye which ensures a high contact surface on the pin without weakening the piston. This may be trapezoidal (narrow at the top, wide at the bottom) or stepped (narrow shoulder at the top, wide at the bottom).

The connecting rod is designed in line with the calculation of the gas and inertial forces (see Crankshaft-assembly design). The case of buckling is also calculated. Both buckling in the plane perpendicular to the crankshaft as a hinged beam and buckling in the crankshaft-axis plane (engine longitudinal plane) as a fixed beam must be checked. The critical cross-sections in the transition from the dimensionally stable connecting-rod eye to the double-T shank are usually recalculated with the aid of FEM simulations (finite-element method). The loads caused by the gripping of the bearing caps must not be underestimated and ignored.

Figure 5: Crankshaft for a commercial-vehicle engine
1 Front end of crankshaft,
2 Main bearing,
3 Connecting-rod bearing journal with oil bore,
4 Main web (without counterweight),
5 Main webs,
6 Intermediate web,
7 Offset connecting-rod bearing journal (split-pin),
8 Counterweight (bolted).



Crankshaft

The crankshaft with its throws (Figure 5) converts the oscillating motion of the piston by way of the connecting rod into a rotatory motion. The translational energy is transformed into a rotating torque, which can be picked off by a coupling.

The main bearings of the crankshaft are fixed in the crankcase, while the crank pins in the throws run in the bearings fixed to the connecting rods. The counterparts to the throws with crank pins are the counterweights, which are either cast, forged or bolted on. Depending on the engine type, the crankshaft main bearings are arranged after each cylinder – particularly in in-line engines and diesel and gasoline engines subject to high loads – or only after every second throw (gasoline engines subject to low loads). V-engines usually feature one crank pin for two adjacent connecting rods on one throw. This also applies to forked connecting rods, which permit two exactly opposing cylinder banks. In order to facilitate a uniform ignition interval in V6 engines with a 90° bank angle, split-pin designs which permit a 30° offset of two crank pins in a very tight space without a main bearing in between are increasingly being used. This form of design is used even in commercial-vehicle diesel engines subject to high loads.

Crankshafts are usually forged (drop forgings or open-die forgings, depending on the size) and, in lower-load applications, increasingly cast (nodular graphite iron). Built-up crankshafts are also used in large engines. Bores for supplying lube oil to the connecting-rod bearings are incorporated in such a way as not to generate further increases in stress.

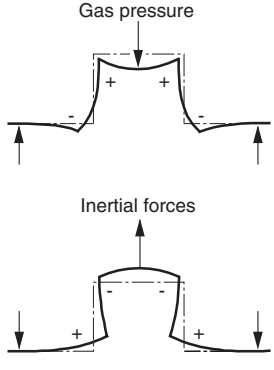
The gas and inertial forces (see Crankshaft-assembly design) subject the crankshaft to bending stress in particular (Figure 6). If a main bearing is incorporated after each throw, the bending loads on the crankshaft are low on account of tangential forces. However, particular attention must be paid to the transition between crank pin and web because this is where the greatest stresses occur. Because of the large number of main bearings, the clamping of the crankshaft represents theoreti-

cally a statically overdefined system – in the simulation, however, a statically defined state which correctly describes the critical state is assumed.

Another important factor is the torsional vibrations, which, on account of the alternating forces of each individual cylinder (torsional force, see Crankshaft-assembly design) excite the overall crankshaft-assembly system and, in view of the not inconsiderable length of the crankshaft, can approach the crankshaft's natural resonance.

The oscillatory system – consisting of piston, connecting rod and crankshaft – is a complex system of constantly changing torsional forces. The further the counterweights are arranged to the outside, the higher the excitation. For this reason, heavy-metal counterweights are also bolted on in some cases. A simple simulation can be performed by assuming a smooth, inertialess and elastic shaft with attached equivalent masses. The oscillation-equation system can then be set up with the aid of the torsion-spring rate of the torque shaft. Cast shafts are ascribed a certain level of intrinsic damping. The crankshaft is thus subjected to both a locally different and a time-varying torsional load.

Figure 6: Crankshaft throw
Primary stresses and deformations due to gas pressure and inertial forces.



The torsional vibrations of the crankshaft, which are also called 2nd-order torsional vibrations (see Crankshaft vibration damper) must be reduced by suitable vibration dampers. This is particularly necessary to protect the crankshaft against excessively high loads.

Crankcase

The crankcase supports the force-transfer mechanism between cylinder head and crankshaft assembly. It bears the crankshaft assembly's support bearings, and incorporates (or holds) the cylinder sleeves. Also included in the block are a separate water jacket and sealed oil chambers and galleries. Furthermore, it must absorb the torque reaction forces and provide locations for mounting many add-on components.

A rigid framework-like structure (support-beam concept) is normally used to absorb the forces; this structure is designed to ensure a direct, linear and bending-moment-free flow of force. This is usually achieved by walls with reinforcing ribs and in the case of the bolt-through design by through-bolts.

The cylinder head is usually bolted as a separate component – in larger engines also as an individual cylinder head – to the crankcase. Often four bolts per head are used, with adjacent cylinders usually sharing two bolts. Six or eight bolts may also be used in engines with high peak pressures, typically above 200 bar. The purpose of this in particular is to distribute as evenly as possible the pressure force of the cylinder-head gasket.

The crankshaft-bearing caps are usually bolted from underneath to the crankcase. In the bolt-through design a long bolt performs this function from the head to the bearing cap. The advantage of this design is that it permits the use of high-tensile steels, which can transmit tensile load better than casting materials.

The crankcase is usually drawn down at the sides to deep below the crankshaft bearings in order to provide sufficient rigidity. But sometimes a separate frame ("bed plate") may also be used which is designed to reduce weight and increase rigidity. Sometimes the side walls are additionally bolted horizontally at the sides

in the crankshaft-bearing area in order to achieve even greater rigidity. Bolted underneath is the oil pan, which can sometimes have a similarly bracing effect. However, this only applies to cast oil pans.

The piston bearing surface can be designed as a separate component (liner), particularly when replacement is not uncommon during the crankcase's service life, for example in commercial vehicles with high mileage ("wet liner" because coolant flows around it). However, the piston bearing surface can also be cast into the crankcase – particularly in cast-aluminum designs – or directly as a bearing surface in gray cast iron with strip hardeners.

Today's crankcases have attained a very high level of function integration and are therefore highly complex components. Frequently the oil-pump housing is also integrated in the casting, as are oil-cooler functions, the coolant-thermostat housing, the balancer-shaft locators, and all the brackets for the accessories.

Aside from the crankshaft, the crankcase is the component which contributes most to the engine weight. Many efforts are therefore made to reduce the weight by using materials other than the customary gray cast iron. These include aluminum materials, which, because of their thermally limited service life, are primarily used in smaller engines, particularly in passenger-car applications. Die casting has become an accepted cost-saving process for manufacturing aluminum, while sand casting is otherwise the regular process used. An "open-deck" construction is used to manufacture diecast crankcases efficiently, i.e. the joint face with the cylinder head features large pass-through cross-sections for the water jacket and the oil return. When cast-iron materials are used, on the other hand, the upper crankcase is closed to make the structure as rigid as possible.

Efforts are also being made to manufacture crankcases from magnesium in order to further reduce weight – but this comes at the expense of a high manufacturing cost. For this reason, a "hybrid" material concept is used where the bearing functions are performed by aluminum and the shielding functions by magnesium inserts. Further increases in strength are provided by other cast-iron variants, compacted graphite iron (CGI) being one such variant which has been more widely used in recent years. Special heat treatment and specific cooling can be used to create in cast iron high-strength (and hard-to-machine) spheroidal graphite instead of the customary lamellar graphite. This makes it possible to produce thinner walls and higher strength with reduced weight.

The crankcase is assigned the most important function of noise emission. Because the crankcase is often joined directly to the combustion chamber, absorbs the high periodic gas and inertial forces and has a relatively large surface, it is extremely important for this surface to be rigid. The standard design methods used to provide high rigidity are ribbing and barreling, which also help to prevent a membrane effect.

Variable compression ratio

A variable compression ratio is desirable both in gasoline engines to reduce the knock tendency close to full load and in diesel engines to reduce the peak pressure at full load. However, up to now no mechanical or hydraulic system has successfully managed to provide this variability in diesel engines. The reason for this is the high loads to which diesel engines are subjected on account of the higher peak pressures and higher pressure gradients compared with gasoline engines.

The designs for variable compression ratio (VCR) can be categorized as follows:

- Solutions which are integrated into the piston (hydraulic actuation of a piston upper section that moves axially in relation to a fixed lower section)
- Solutions which alter the length of the connecting rod (sliding solution, articulated connecting rod, eccentric connecting-rod bearing in the small or big connecting-rod eye)
- Formulations for an altered crankshaft position relative to the cylinder head (swiveling or axial movement of the crankcase, eccentric crankshaft bearing)
- Coupling lever for varying the effective piston stroke (multilink, crankshaft-drive lever system, displacement alteration with ε variation)

– Additional volumes which are increased or decreased (variable additional volume in the cylinder head)

Formulations for variable compression ratios in gasoline engines cover a range of approx. $\varepsilon = 7$ to 14.

Cylinder head

The main functions of the cylinder head are to supply the cylinder with fresh mixture, to discharge the exhaust gas, and to absorb the cylinder pressure forces.

Cylinder head

The cylinder head seals off the upper end of the crankcase and the cylinder barrel. It houses the gas-exchange valves and the spark plugs in gasoline engines, the fuel injectors in direct-injection gasoline engines, and the fuel injectors and possibly the sheathed-element glow plugs in diesel engines. Together with the piston, it also provides the desired combustion-chamber shape. In the vast majority of passenger-car engines, the entire valve gear is also mounted in the cylinder head.

The cylinder head performs a large number of functions and exhibits complex load profiles. The challenge faced here, as well as precisely controlling gas exchange and charge movement, which directly influence engine emissions, efficiency and power output, is above all to cool the hot side of the cylinder head and maintain control over the complex mechanical and thermal loads involved.

Based on the gas-exchange concepts, a distinction is made between two basic design configurations. In the counterflow cylinder head intake and exhaust passages open into the same side of the cylinder head (Figure 7b). This limits the space available for the intake and exhaust-gas passages, but due to the short flow tracts, this represents a substantial advantage in supercharged applications without intercooling. This design, with the gas supply and discharge tracts on a single side, also provides practical advantages in transverse-mounted engines.

In the crossflow cylinder head intake and exhaust passages are located on opposite sides of the engine (Figure 7a), providing a diagonal flow pattern for the intake and exhaust gases. This layout's advantages include more freedom in intake and exhaust-tract design as well as less complicated sealing arrangements. Typically, vehicle-specific boundary conditions are crucial to determining whether a counterflow or crossflow cylinder head is used.

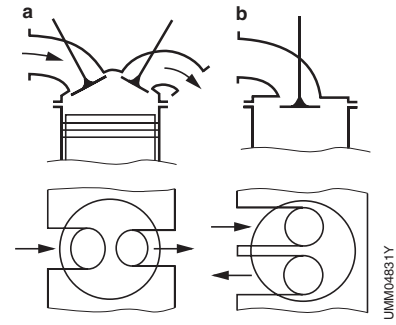
This decision also determines the valve arrangement used. Counterflow applications require a twisted valve arrangement in today's typical multi-valve designs. The main advantage of a twisted valve arrangement is that it is a simple way of delivering high cylinder-whirl values. A parallel valve arrangement requires greater efforts at the passage end to deliver a high cylinder whirl (Figure 8).

Today's passenger-car applications use almost exclusively a continuous single-piece cylinder head. In V-engines one cylinder head per cylinder bank is used. Individual cylinder heads are frequently still used in commercial-vehicle engines, which sometimes still feature a block-mounted camshaft with a push-rod assembly.

In contrast to passenger-car applications, in which the cylinder head is often made from aluminum, which as well as weight advantages improves heat transfer from the combustion chamber, commercial-vehicle applications utilize gray-cast-iron designs in view of the fact that

Figure 7: Cylinder-head identification according to intake and exhaust-tract location

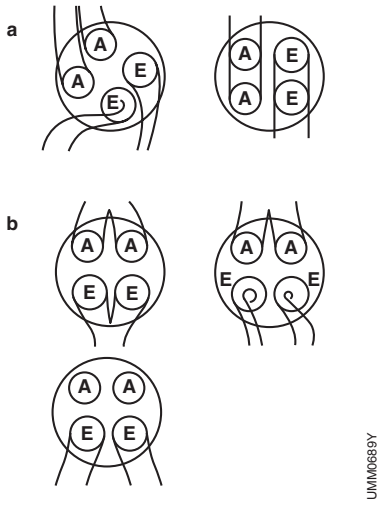
a) Crossflow design,
b) Counterflow design.



resistance to fatigue cannot be achieved by light alloys at hight peak pressures and with large cylinder diameters.

The cylinder head is subject to very high loads. On the one hand, frequently varying temperatures and the accompanying expansions give rise to stresses which can only be absorbed by appropriate design measures (individual cylinder head, suitable cavities in a multi-cylinder head). Thermomechanical strength is determined by the number of cold-hot load profiles.

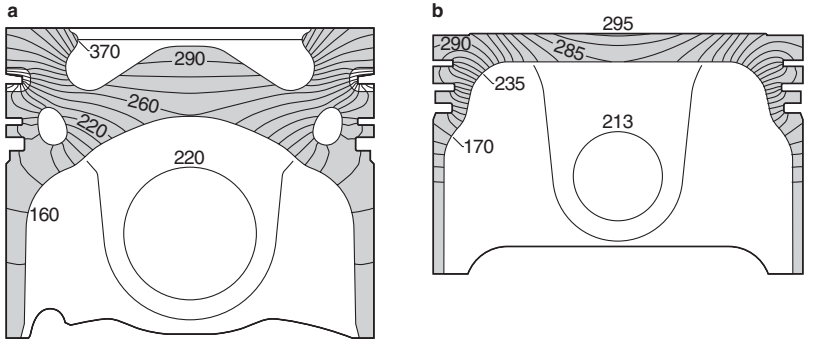
Figure 8: Valve arrangements
a) Twisted valve arrangements,
b) Parallel valve arrangements.
E Intake, A Exhaust.



In multi-valve designs the land between the exhaust valves is subjected to particularly high temperatures. At full load slightly critical temperatures above 350 °C are reached here. Similarly high temperatures are encountered in the recess-edge area (Figure 9). Gray-cast-iron designs are especially prone to the risk of hot-gas corrosion, an oxidation of the surface which can culminate in component-jeopardizing cracking.

The cylinder-head gasket, which is located between the crankcase and the cylinder head, performs an important function. Firstly, it must seal the combustion chamber even in the case of high cylinder pressures and, secondly, it must ensure the flow of engine oil and cooling water between crankcase and cylinder head. Typically, metal materials are used to seal the combustion chamber; these metals adapt to the surface by means of plastic deformation and establish the seal through their elasticity. Elastomer gaskets or geometrically optimized metal-sheet steel gaskets are used.

Figure 9: Piston operating temperatures in motor-vehicle engines at full load (schematic), values in °C
a) Passenger-car diesel-engine piston, 16 MPa ignition pressure, 58 kW/*l*,
b) Passenger-car gasoline-engine piston, 7.3 MPa ignition pressure, 53 kW/*l*.



To provide a uniform downward pressure on the cylinder-head gasket, it is possible to increase the number of cylinder-head bolts to such an extent that already commercial-vehicle engines with eight bolts per cylinder head are in use.

Engines without cylinder-head gaskets are the exception and require the crankcase deck and the underside of the cylinder head to be finished to very high levels of accuracy.

Valve gear

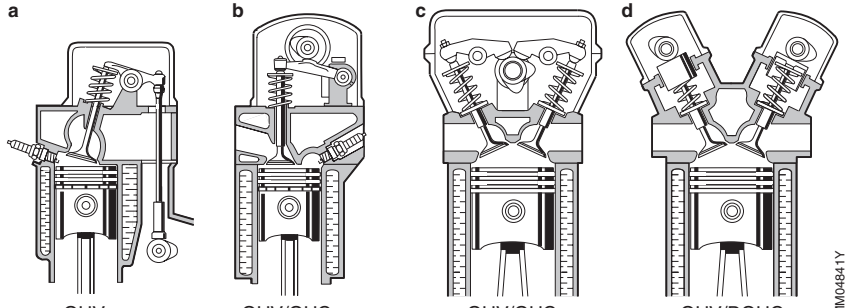
It is the function of the valve-gear assembly in a 4-stroke engine to permit and to control the exchange of gases in the internal-combustion engine. The valve gear includes the intake and exhaust valves, the springs which close them, the camshaft drive assembly, and the various force-transfer devices (Figure 10). In various widely-used designs, the camshaft is located in the cylinder head.

When the camshaft is installed within the block, the rocker arm is actuated not directly by the cam lobe, but by the intermediate action of a push rod and a lifter (push-rod assembly, Figure 10a).

In the case of the cam-follower or single rocker-arm assembly actuated by an overhead cam, the cam lobe's lateral and linear forces are absorbed and relayed by a cylinder-head mounted lever rocking back and forth between the cam lobe and the valve. In addition to transferring forces and absorbing lateral forces, the intermediate rocker arm can also be designed to magnify the cam pitch effect (Figure 10b).

In the case of the twin rocker-arm assembly actuated by an overhead cam, the rocker arm acts as the force-transfer element with its tilt axis located between the camshaft and the valve. Here, too, the rocker arm is usually designed as a cam-

Figure 10: Valve-timing designs
(Source: [1]).
a) Push-rod assembly, b) Single rocker-arm assembly,
c) Twin rocker-arm assembly, d) Overhead bucket-tappet assembly.
OHV Overhead valves,
OHC Overhead camshaft,
DOHC Double overhead camshaft.



lift multiplier to produce the desired valve travel (Figure 10c).

In the case of the overhead bucket-tappet assembly, a "bucket" moving back and forth in the cylinder head absorbs the cam lobe's lateral force, while transferring its linear actuating pressure to the valve stem (Figure 10d).

Valve arrangements

The valve-control arrangement and the design of the combustion chamber are closely interrelated. Today, nearly all valve assemblies are overhead units mounted in the cylinder head. In diesels and simpler spark-ignition (gasoline) engines, the valves are parallel to the cylinder axis, and are usually actuated by twin rocker arms, bucket tappets or single rocker arms. With increasing frequency, current spark-ignition (SI) engines designed for higher specific outputs tend to feature intake and exhaust valves which are inclined towards each other. This configuration allows larger valve diameters for a given cylinder bore, while also providing greater freedom for optimizing intake and exhaust passage design. Twin rocker-arm assemblies actuated by overhead cams are used most often here. High-performance and racing engines are increasingly using four valves per cylinder and overhead bucket-tappet valve assemblies.

The valve-timing diagram (Figure 11) for an engine shows the opening and closing times of the valves, the valve-lift curve, the maximum valve lift, and the valve velocity and acceleration.

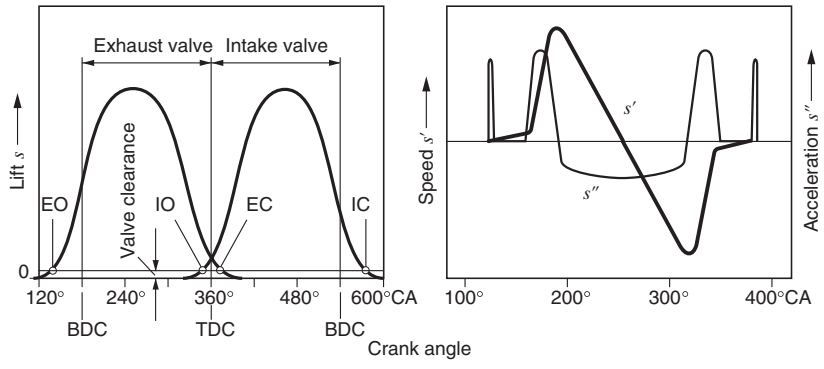
Typical valve-acceleration rates for passenger-car OHC (overhead camshaft) valve assemblies:
s'' = 60…65 mm (b/ω^2) → 6,400 m/s^2 at 6,000 rpm for single and twin rocker-arm assemblies.
s'' = 70…80 mm (b/ω^2) → 7,900 m/s^2 at 6,000 rpm for overhead bucket-tappet assemblies.
For heavy-duty commercial-vehicle engines with block-mounted camshafts:
s'' = 100…120 mm (b/ω^2) → 2,000 m/s^2 at 2,400 rpm.

Valve, valve guide and valve seat

The materials employed in manufacturing valves are heat and scale-resistant. The valve seat's contact surface is frequently hardened. Another proven method for improving the thermal-conductivity characteristics of exhaust valves is to fill their stems with sodium. To extend service life and improve sealing, valve-rotating systems (rotocaps) are also in common use.

The valve guides in high-performance engines must feature high thermal con-

Figure 11: Valve-timing diagram
s Valve lift, s' Valve velocity, s'' Valve acceleration.
IC Intake closes, IO Intake opens, EC Exhaust closes, EO Exhaust opens,
BDC Bottom dead center, TDC Top dead center.



ductivity and good antifriction properties. They are usually pressed into the cylinder head and are often supplemented by valve-stem seals at their cold ends for reducing oil consumption.

Valve-seat wear is generally reduced by making the valve-seat inserts of cast or sintered materials and shrink-fitting them into the cylinder head.

Cam-lobe design and timing dynamics

The cam lobe must be able to open (and close) the valve as far, as fast and as smoothly as possible. The closing force for the valves is applied by the valve springs, which are also responsible for maintaining contact between the cam lobe and the valve. Dynamic forces impose limits on cam and valve lift.

Rarely encountered are designs which use their own mechanism (Desmotronic) to restore the valves with the advantage of increased valve velocities and thus lift curves which more closely resemble the ideal rectangular case. Conventional springs must always ensure a frictional connection between camshaft, intermediate elements and valve while adhering to the permissible surface load. A slight clearance can be encountered only when cam lift is not active. Modern engines operate here with hydraulic valve-clearance compensation. Otherwise valve clearance must be regularly checked; measurements of 0.1 to 0.2 mm are common at the intake and exhaust valves.

Camshaft drive

To drive the valve gear, the camshaft must be connected to the crankshaft in such a way that one crankshaft revolution results in a one half camshaft revolution in a 4-stroke engine.

For passenger-car applications with overhead camshaft(s) either timing-chain or toothed-belt drives are used in modern engines. Both drives require a tensioning device, which acts on the free running length, the loose side, to prevent uncontrolled vibration.

In view of the fully utilized minimum space between piston and valve in modern engines, it is very important for the camshaft drive to operate safely and reliably, because in the event of damage contact can have serious consequences for the valve gear.

For this reason, the pressure pins for the tensioning rails in a chain drive or the toothed belt are wearing parts which must be replaced on a regular basis. Modern duplex chains are highly durable and are only subject to minimal wear. The drawbacks of a timing chain are its need to be lubricated and its elongation during operation. Some major advantages, however, are that it cannot jump off the gear wheel and it is not subject to the risk of cracking.

Transmission by gear wheels in a geartrain is typically used in commercial-vehicle engines or engines with block-mounted camshafts. In engines with overhead camshafts in particular, this is a very expensive solution, but one which provides for extending engine running times on account of the reliable transmission.

Expensive vertical shafts, whose axes run at right angles to the crankshaft and camshaft, are rarely used.

Oil supply

As well as lubricating the tribologically critical pairings of the crankshaft drive, the cylinder head and other components, the oil supply serves to remove local contaminants, combustion residues and wear particles which are filtered out in the oil-filter unit. Additional functions include the dissipation of heat in areas subject to thermal load, such as the friction bearings in the crankshaft drive or for oil-cooled pistons, and damping of vibrations in bearings.

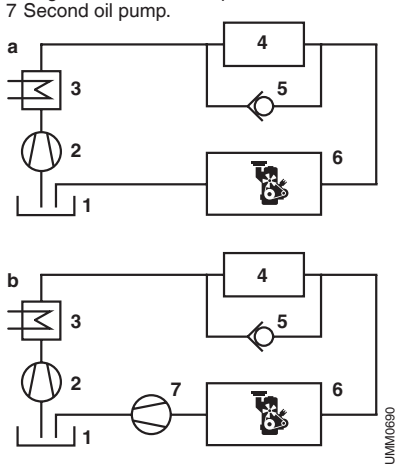
With the system of forced-feed lubrication conventionally used (see also "Engine lubrication"), the oil pump, a gear pump that typically operates as a positive-displacement pump, delivers a defined volumetric flow from the oil sump through an oil-filter unit (Figure 12a). For safety reasons, the oil-filter unit is often equipped with a bypass valve and a pressure valve in a full-flow design. The oil cooler provided in engines subject to higher loads is cooled by either air or coolant. The engine oil flows though oil ducts and by the force of gravity back into the oil sump and into the oil pan typically located underneath the crankcase. That is why forced-feed lubrication is also known as wet-sump lubrication. In addition to the direct supply of oil by the oil pump, the rotational motion of the crankshaft effects a fine conditioning of the oil mist in the crankcase.

The opposite system to conventional wet-sump lubrication is the more expensive dry-sump lubrication, because it requires a second oil pump to remove the oil from the engine compartment (Figure 12b). The advantages are a constantly guaranteed lube-oil supply under conditions of high lateral acceleration or inclination and freedom in deciding on where to locate the oil-supply system. These give rise to lower engine heights and a possible increase in the engine-oil quantity, which is beneficial to engine cooling.

The mixture and total-loss lubrication frequently used in 2-stroke and Wankel engines is no longer relevant to present-day automobile applications.

Figure 12: Oil supply

a) Forced-feed lubrication (wet-sump lubrication),
b) Dry-sump lubrication.
1 Oil reservoir, 2 Oil pump, 3 Oil cooler,
4 Filter unit, 5 Bypass valve,
6 Engine with lubrication points,
7 Second oil pump.



Cooling

In order to avoid thermal overload, combustion of the lubricating oil on the piston's sliding surface, and uncontrolled combustion due to excessive component temperatures, the components surrounding the hot combustion chamber (cylinder liner, cylinder head, valves and in some cases the pistons themselves) must be intensively cooled (see also "Engine cooling").

As water has a high specific heat capacity and provides efficient thermal transition between the materials, most contemporary vehicle engines are water-cooled. Moreover, it is possible, under very hot local boundary conditions when the cooling medium is locally evaporated, to achieve through the heat of evaporation an additional cooling effect which together with the condensation behavior in colder adjacent areas reduces unwanted high local temperature gradients.

Air/water recirculation cooling is the most prevalent system (Figure 13). It comprises a closed circuit allowing the use of anticorrosion and antifreeze additives. The coolant is pumped through the engine and through an air/water radiator. The cooling air flows through the radiator in response to vehicle movement and/or

is forced through it by a fan. The coolant temperature is regulated by a thermostatic valve which bypasses the radiator as required.

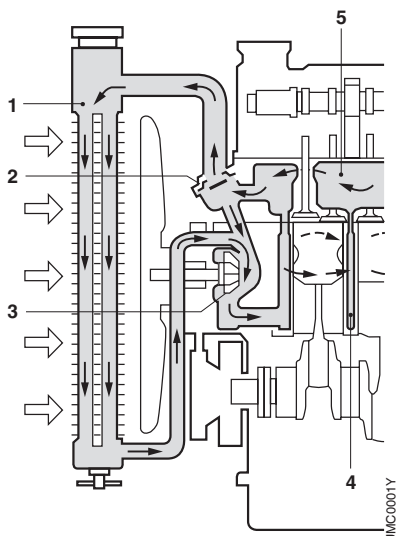
Air-cooled concepts still play a subordinate role today. Modern emission regulations call for effective engine cooling with simultaneously high specific power output. This can practically no longer be achieved with air cooling, a system which is indeed robust and low-maintenance, but is also not as effective. A crucial drawback of air-cooled concepts is their acoustic performance. The cooling fins act like a resonator and therefore result in a marked increase in sound emissions.

Crankshaft vibration damper

The crankshaft is subjected to high-frequency vibrations (2nd-order torsional vibrations), particularly in the region of resonant frequencies. Vibration dampers are therefore used to moderate dangerous crankshaft resonance phenomena.

Figure 13: Water cooling system with coolant circuit

1 Radiator,
2 Thermostat,
3 Water pump,
4 Water passages in cylinder block,
5 Coolant passages in cylinder head.



The choice of vibration-damper design ranges from simple flywheel masses, which through elastically damping layers – e.g. rubber – provide a countervibration system, through more sophisticated oil-damped flywheel masses, in which the viscosity of the oil and friction surfaces provide the damping effect, up to complex pendulum adsorbers. The vibration damper is generally mounted at the free crankshaft end (front).

1st-order torsional vibrations, i.e. vibrations resulting from ignition and inertial forces, which can result in excitation of transmission and input-shaft vibrations, are reduced by dual-mass flywheels or other damping technologies. Phenomena such as clutch grabbing, transmission chatter and starting jerks can be optimized by such vibration absorbers, since all the following components – clutch, transmission and input shafts – likewise represent an oscillatory system. In a dual-mass flywheel the two flywheel masses are coupled to each other by means of springs. Torsional-vibration dampers are also commonly featured in clutches – both components must naturally be tuned together.

Design and tuning are difficult to simulate and must in the final analysis be conducted in tests, both in the vehicle with the entire drivetrain (1st order: low-frequency drivetrain resonance phenomena) and on the test bench (2nd order: high-frequency crankshaft resonances).

Reciprocating-piston engine types

The internal-combustion engine is characterized by its great variability and application-specific adaptability. In particular, the arrangement of the individual cylinders provides for numerous variants.

Arrangements

In principle, a huge number of possible cylinder arrangements is conceivable. Few of these arrangements have proven to be particularly effective for automotive applications (Figure 14).

Radial engines are not suitable for automotive applications on account of their height and are therefore not used.

It is important to note the difference between the opposed-cylinder (boxer) engine and the V-engine with 180° included angle between cylinder banks. In the boxer engine the pistons always move in opposite directions, which is why the inertial forces cancel each other out. In the V-engine, on the other hand, the pistons of two opposed cylinders always move in the same direction.

It is also important to note the design of V- and W-engines, in which the connecting-rod/crankshaft link (large connecting-rod eye) of associated cylinders is effected at slightly offset crank pins.

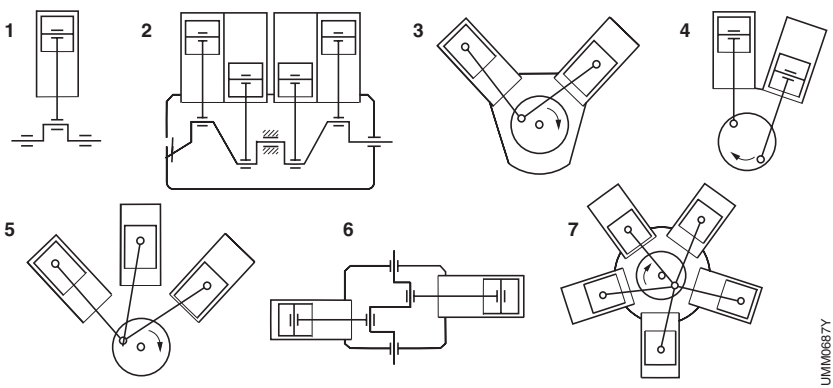
In multi-piston units compression is typically generated by several working pistons. In the U-engine the pistons undertake an approximately parallel motion. In the opposed-piston engine the pistons undertake a typically opposed motion. Engine variants with more than one piston per combustion chamber have not gained acceptance. Weight and engine size are not suitable for automotive applications.

Definitions

The following definitions in accordance with DIN 73021 [2] as viewed looking at the end opposite the power-output end apply only to motor-vehicle engines. For internal-combustion engines for general and marine use, the reverse direction – i.e. as viewed looking at the power-output end – is standardized (ISO 1204 [3]).

Figure 14: Reciprocating-piston engine types

1 Single cylinder (individual cylinder),
2 In-line engine (arrangement of individual cylinders in line),
3 V-engine (arrangement of individual cylinders in two banks in a V-shape),
4 VR-engine (arrangement of individual cylinders in two banks in a V-shape, where the two banks are particularly close to one another),
5 W-engine (arrangement of individual cylinders in three banks in a W-shape),
6 Opposed-cylinder (boxer) engine (arrangement of individual banks opposing one another),
7 Radial engine (radial arrangement in one or more planes).



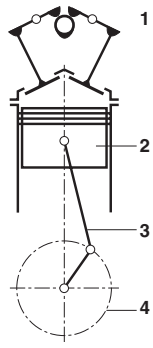
Clockwise rotation
The direction of rotation is clockwise, as viewed looking at the end opposite the power-output end.

Counterclockwise rotation
The direction of rotation is counterclockwise, as viewed looking at the end opposite the power-output end.

Numbering the cylinders
The cylinders are numbered consecutively 1, 2, 3, etc. in the order in which they would be intersected by an imaginary reference plane, as viewed looking at the end of the engine opposite the power-output end. This plane is located horizontally to the left when numbering begins; the numerical assignments then proceed clockwise about the longitudinal axis of the engine. If there is more than one cylinder in a reference plane, the cylinder nearest the observer is assigned the number 1, with consecutive numbers being assigned to the following cylinders. Cylinder 1 is to be identified by the number 1.

Firing sequence
The firing sequence is the sequence in which combustion is initiated in the cylinders. Engine design configuration, uniformity in ignition intervals, ease of crankshaft manufacture, optimal crankshaft load patterns, etc. all play a role in defining the firing sequence.

Figure 15: Crankshaft assembly for reciprocating-piston engine (principle)
1 Valve gear,
2 Piston,
3 Connecting rod,
4 Crankshaft.



Crankshaft-assembly design

Crankshaft-drive kinematics
The kinematics of the single-cylinder crankshaft assembly can be determined from the geometrical assignments of piston and piston-pin axis, connecting rod and crankshaft (crankshaft radius equal to one half stroke) (Figure 15). If piston travel x at top dead center is assumed to be zero, the following is obtained with crank radius r and connecting-rod length l (Figure 16):

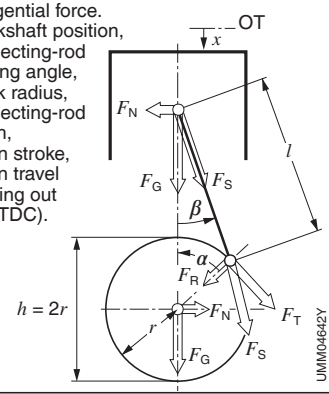
$$x = r\,(1 - \cos\alpha) + l\,(1 - \cos\beta)\ .$$

Where
$r \sin\alpha = l \sin\beta$
and
$\lambda = r/l$
the following is obtained:

$$x = r\left(1 - \cos\alpha + \frac{1}{\lambda}\left(1 - \sqrt{1 - \lambda^2 \sin^2\alpha}\right)\right)$$

An offset of the piston pin is performed by some manufacturers. Noise and friction advantages can be expected on account of the changing piston location and depending on the position of the connecting rod. The offset can be effected either in the piston by shifting the piston pin from the center position or by offsetting the crankshaft.

Figure 16: Gas-force components shown on a basic crankshaft drive
F_G Piston-pin force (gas force),
F_S Connecting-rod force, F_R Radial force,
F_N Piston normal force,
F_T Tangential force.
α Crankshaft position,
β Connecting-rod pivoting angle,
r Crank radius,
l Connecting-rod length,
h Piston stroke,
x Piston travel (starting out from TDC).



If the offset for positive crank angles is assumed to be positive and the quantity

$$\delta = \frac{\text{Offset}}{\text{Connecting-rod length}}$$

is introduced, this leads to the following function of the piston travel:

$$x = r\left(1 - \cos\alpha + \frac{1}{\lambda}\left(1 - \sqrt{1 - (\lambda \sin\alpha - \delta)^2}\right)\right)$$

Figure 17 shows by way of example the influence of the stroke/connecting-rod ratio and the offset. However, the differences with normal offset values in the millimeter range ($\delta < 0.04$) are clearly smaller.

By developing the root function into a Taylor series (around $x = 0$: Maclaurin series) and replacing the powers of the trigonometric functions with the multi-harmonic functions, the following expression can be obtained [4]:

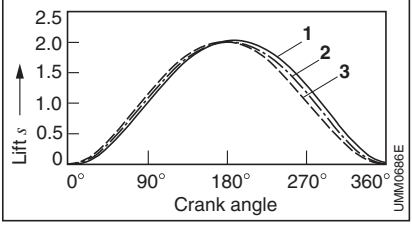
$$x = r\,[1 + \tfrac{1}{4}\lambda + \tfrac{3}{64}\lambda^3 + \ldots - \cos\alpha - (\tfrac{1}{4}\lambda + \tfrac{3}{64}\lambda^3 + \ldots)\cos 2\alpha + (\tfrac{3}{64}\lambda^3 + \ldots)\cos 4\alpha + \ldots]\,.$$

This expression shows that there are higher harmonics on account of the crankshaft-drive kinematics, and these are also called engine orders (multiples of the engine speed).

Because normal values for λ are around 0.3, the higher-level λ terms are readily ignored and the following simplified function is assumed for further calculations:

$$x = r\,[1 + \tfrac{1}{4}\lambda - \cos\alpha - \tfrac{1}{4}\lambda\cos 2\alpha\,]\,.$$

Figure 17: Piston-travel function
1 $\lambda = \infty$,
2 $\lambda = 0.3$,
3 $\lambda = 0.3$, $\delta = 0.1$.



This simplification may not be used if detailed vibration analyses and resonance phenomena are to be examined.

From the simplified equation the following piston speeds v and accelerations a are obtained, where the angular velocity $d\alpha/dt = \omega = 2\pi n$ (n speed) was introduced:

$$v = r\omega\,(\sin\alpha + \frac{\lambda}{2}\sin 2\alpha)\,,$$

$$a = r\omega^2\,(\cos\alpha + \lambda\cos 2\alpha)\,.$$

Here too there are higher harmonics (orders), which must not be disregarded when resonances are examined.

Crankshaft-assembly kinetics

The forces acting on the crankshaft assembly and the resulting moments can initially be derived without inertial forces as follows (Figure 16).

The piston-pin force results as gas force from the combustion-chamber pressure and the piston area. The following applies:

$$F_G = (p - p_{KGH})\,A_{piston}\,.$$

The connecting-rod force is obtained from the vectorial analysis of the piston-pin force in the direction of the connecting rod. The following applies:

$$F_S = \frac{F_G}{\cos\beta} = \frac{F_G}{\sqrt{1 - \lambda^2\sin^2\alpha}}$$

The piston normal force F_N is the vectorial component of the piston-pin force normal to the cylinder wall and for balancing the connecting-rod force:

$$F_N = F_G \tan\beta = \frac{F_G\,\lambda\,\sin\alpha}{\sqrt{1 - \lambda^2\sin^2\alpha}}\,.$$

This force contributes significantly to the friction of the piston on the cylinder barrel. The side which is touched by the piston after top dead center on account of the combustion pressure is termed the major thrust face, the opposite side the minor thrust face. The greatest friction is therefore created shortly after TDC on the major thrust face.

The tangential force at the crankshaft crank pin contributes to an acceleration of the crankshaft and thus to the build-up of torque at the crankshaft. It is created from the vectorial analysis of the connecting-rod force.

$$F_T = \frac{F_G \sin(\alpha+\beta)}{\cos\beta} = F_G\left(\sin\alpha + \frac{\lambda}{2}\sin 2\alpha/\sqrt{1-\lambda^2\sin^2\alpha}\right).$$

Here too the radicand can be simplified in a series expansion as

$$F_T \approx F_G\,(\sin\alpha + \frac{\lambda}{2}\sin 2\alpha).$$

The radial force F_R at the crankshaft crank pin is:

$$F_R = F_G\frac{\cos(\alpha+\beta)}{\cos\beta} = F_G\left(\cos\alpha + \lambda\frac{\sin^2\alpha}{\sqrt{1-\lambda^2\sin^2\alpha}}\right)$$

or approximately

$$F_R \approx F_G\,(\cos\alpha - \frac{\lambda}{2} + \frac{\lambda}{2}\cos 2\alpha).$$

The inertial forces can be split into oscillating and rotating components. The piston, piston-ring and piston-pin weights m_K belong to the oscillating component and can be punctually assigned to the piston pin.

The crankshaft web with crank pin belong to the rotating component. Here the mass is usually reduced to the crank radius and punctually assigned to the crank-pin center point. The following applies:

$$m_W = \Sigma\,\frac{m_i\,r_{si}}{r},$$

where m_i is the respective mass component of web, pin, etc. and r_{si} is the corresponding center-of-mass radius.

Because of the connecting rod's swinging motion, it is useful to split the connecting-rod mass into an oscillating component and a rotating component. This is done exactly where the connecting-rod-mass center of gravity and mass moment of inertia are known by assuming two dynamically equal individual masses at the small and big connecting-rod eyes and by calculating the equilibrium formulation of force, moment and mass inertia. Usually one third of the connecting-rod mass m_{Pl} can be assumed to be oscillating and two thirds to be rotating. Then, with $m_o = m_K + 1/3\ m_{Pl}$ as the oscillating mass and the piston acceleration (see below), an oscillating inertial force is obtained at:

$$F_o \approx m_o\,r\,\omega^2\,(\cos\alpha + \lambda\cos 2\alpha).$$

The oscillating inertial force therefore increases quadratically with the engine speed ($\omega = 2\pi n$) and has a first-order component and a lower second-order component.

The rotating inertial force is obtained as centrifugal force from the reduced mass $m_r = m_W + 2/3\ m_{Pl}$ and the rotation speed at:

$$F_r = m_r\,r\,\omega^2.$$

The rotating inertial force likewise increases with the square of the speed, but has not higher orders. The rotating inertial force can therefore be easily balanced by counterweights which rotate at engine speed. The rotational irregularities of the crankshaft are so small compared with these forces that they can be ignored in the balancing of masses.

As shown in the crankshaft-assembly kinematics, higher harmonics – higher engine orders – occur on account of the crankshaft-drive geometry. Aside from the 1st and 2nd engine orders, however, the amount of the 4th order and the higher components decreases rapidly, and is generally ignored for a balancing of masses.

Balancing of masses in a single-cylinder engine

The component of rotating masses in a single-cylinder engine can be fully balanced by appropriate counterweights on the crank pin. Weights are normally provided for on both sides and the weights must be balanced merely with the center-of-mass radius. The oscillating forces can be represented with force vectors (Figure 18) when they are modeled rotating in the opposite direction in each case with half the amount.

Therefore two contra-rotating shafts with weights can be used to balance the oscillating inertial forces. The horizontal component then cancels out and at least the first order of the oscillating inertial-force component can be compensated.

A virtually full balancing of masses requires further balancer shafts, which must rotate at twice the engine speed in order to balance completely the second-order component of the oscillating mass component.

A compromise is often reached due to the fact that the expenditure associated with contra-rotating shafts is high and at least with the first order already not inconsiderable weights would have to be realized in the shafts. For example, half the oscillating mass can be integrated in the counterweights. The free inertial forces acting outwards are now reduced in the cylinder longitudinal direction by one half; however, additional transverse forces are now created on account of the excessive rotating components (see Table 2). A half compensation of this kind is termed a 50 % balancing rate. Usual figures are 100 % rotatory and 50 % oscillatory balancing of masses.

Balancing of masses in a multiple-cylinder engine

The inertial forces of a multiple-cylinder engine are composed of the inertial forces of each individual cylinder which are superimposed according to the crankshaft throw. In addition, free moments of inertia are created on account of the cylinder spacing. Table 3 shows a summary of the possible transverse and longitudi-

Figure 18: Complete balancing of 1st- and 2nd-order masses in a single-cylinder assembly

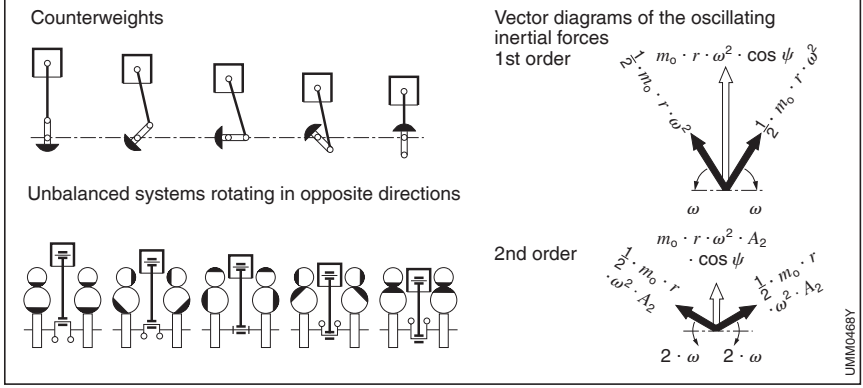
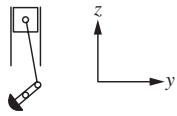


Table 2: Balancing of masses on a single-cylinder engine, dependent on the balancing rate

		Balancing rate		
		0 %	50 %	100 %
Size of counterweight	$m_G \triangleq$	m_r	$m_r + 0.5m_0$	$m_r + m_0$
Residual inertial force (z) 1st order	$F_{1z} =$	$m_0 \cdot r \cdot \omega^2$	$0.5 \cdot m_0 \cdot r \cdot \omega^2$	0
Residual inertial force (y) 1st order	$F_{1y} =$	0	$0.5 \cdot m_0 \cdot r \cdot \omega^2$	$m_0 \cdot r \cdot \omega^2$

nal tilting moments that occur and the free inertial forces.

The mutual balancing of the inertial forces is one of the essential factors determining the selection of the crankshaft's configuration, and with it the design of the engine itself. The inertial forces are balanced if the common center of gravity for all moving crankshaft-assembly components lies at the crankshaft's midpoint, i.e. if the crankshaft is symmetrical (as viewed from the front). This is represented with the 1st-order and 2nd-order star diagrams (Table 4).

The 2nd-order star diagram for the four-cylinder in-line engine is asymmetrical, meaning that this order is characterized by free inertial forces. These forces can be balanced using two countershafts rotating in opposite directions at double the rate of the crankshaft (Lanchester system).

Table 5 shows a summary of free forces and moments based on balancing of masses for different numbers of cylinders and throw variants.

Torsional force

Moving masses are exposed to constantly changing acceleration, thereby generating inertial forces. The cyclically occurring pressure forces in the cylinders are termed gas forces. Both form in total internal and external forces and moments. The internal forces and moments must be

Table 3: Transverse and longitudinal tilting moments, free inertial forces in multiple-cylinder engines

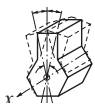
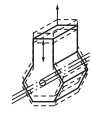
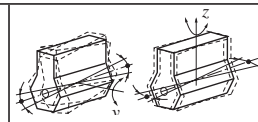
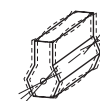
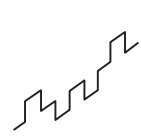
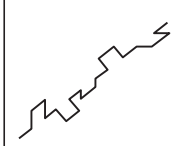
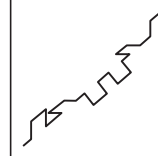
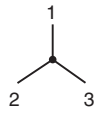
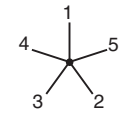
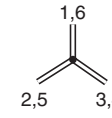
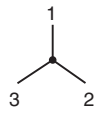
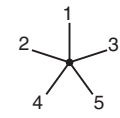
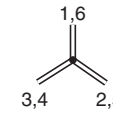
Forces and moments at the engine				 UMM0469Y
Designation	Oscillating torque, transverse tilting moment, reaction torque	Free inertial force	Free inertial moment, longitudinal tilting moment about the y-axis (transverse axis) ("pitching" moment) about the z-axis (vertical axis) ("rolling" moment)	Internal bending moment
Cause	Tangential gas forces as well as tangential inertial forces for the ordinals 1, 2, 3 and 4	Unbalanced oscillating inertial forces 1st order in 1- and 2-cylinders; 2nd order in 1-, 2- ,4-cylinders	Unbalanced oscillating inertial forces as a composite of 1st- and 2nd-order forces	Rotating and oscillating inertial forces
Design factors	Number of cylinders, ignition intervals, displacement, p_i, ε, p_z, m_0, r, ω, λ	Number of cylinders, crank configuration m_0, r, ω, λ	Number of cylinders, crank configuration, cylinder spacing, counterweight size influences inertial-torque components about the y- and z-axes m_0, r, ω, λ, a	Number of throws, crank configuration, engine length, engine-block rigidity
Remedy	Can only be compensated for in exceptional cases	Free mass effects can be eliminated by rotating balancing systems, however, this process is complex and therefore rare; crank sequences with limited or no free mass effects are preferable		Counterweights, rigid engine block
	Shielding of the environment through flexible engine mounts (in particular for orders ≥ 2)			

Table 4: Star diagram for in-line engines

	3-cylinder	4-cylinder	5-cylinder	6-cylinder
Crank sequence				
Star diagram 1st order				
Star diagram 2nd order				

UMM04711Y

absorbed by the components, especially the crankshaft and crankcase, while the external effects act via the engine bearings on the support structure and impart oscillations to the chassis or engine foundation.

If the periodic gas force acting on the piston and the periodic mass inertial forces acting on the piston, connecting rod and crankshaft assembly are grouped together, they generate a sum of tangential force components at the crankshaft journal. When multiplied by the crank radius, this produces a periodically variable torque value.

With multiple-cylinder engines, the tangential-pressure curves for the individual cylinders are superimposed with a phase shift dependent on the number of cylinders, crankshaft-journal configuration, crankshaft design, and the firing sequence. The resulting composite curve is characteristic for the engine design, and covers a full working cycle (i.e. two crankshaft rotations for 4-stroke engines, Figure 18). It can be depicted in a torsional-force diagram. This variable torsional force and the resulting torque generate, depending on the available moment of inertia J, a variable rotation speed ω:

$$\frac{d\omega}{dt} = \frac{M(t)}{J}$$

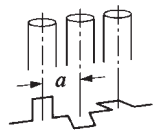
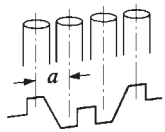
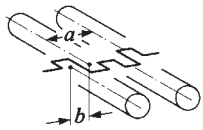
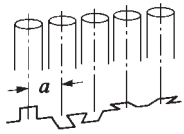
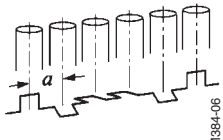
with all the superimposed and newly created engine orders (there are also half orders). This deviation from the speed constant is called the coefficient of cyclic variation and is defined as follows:

$$\delta_S = \frac{\omega_{max} - \omega_{min}}{\omega_{min}}.$$

This coefficient of cyclic variation is reduced to a level suitable for the application by energy-storage mechanisms, such as a flywheel mass for example. The torsional vibrations which can be traced back to the torsional force depicted are also called 1st-order torsional vibrations. These are not to be confused with the (high-frequency) torsional vibrations resulting from elastic deformations and natural resonances of the crankshaft (Table 5), which are referred to as 2nd order.

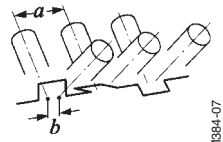
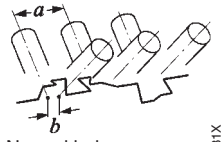
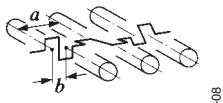
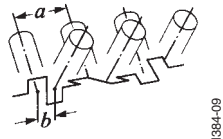
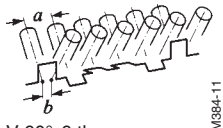
Table 5: Free forces and moments of the 1st and 2nd order, and ignition intervals of the most common engine designs

$F_r = m_r\, r\, \omega^2 \qquad F_1 = m_0\, r\, \omega^2 \cos\alpha \qquad F_2 = m_0\, r\, \omega^2\, \lambda \cos 2\alpha$

Cylinder arrangement	Free forces of 1st order[1]	Free forces of 2nd order	Free moments of 1st order[1]	Free moments of 2nd order	Ignition intervals
3-cylinder					
 VM384-01 In-line, 3 throws	0	0	$\sqrt{3} \cdot F_1 \cdot a$	$\sqrt{3} \cdot F_2 \cdot a$	240°/240°
4-cylinder					
 VM384-02 In-line, 4 throws	0	$4 \cdot F_2$	0	0	180°/180°
 VM384-03 Opposed-cylinder (boxer), 4 throws	0	0	0	$2 \cdot F_2 \cdot b$	180°/180°
5-cylinder					
 VM384-05 In-line, 5 throws	0	0	$0.449 \cdot F_1 \cdot a$	$4.98 \cdot F_2 \cdot a$	144°/144°
6-cylinder					
 VM384-06 In-line, 6 throws	0	0	0	0	120°/120°

[1] Without counterweights.

Table 5 (continued): Free forces and moments of the 1st and 2nd order, and ignition intervals of the most common engine designs

Cylinder arrangement	Free forces of 1st order [1]	Free forces of 2nd order	Free moments of 1st order [1]	Free moments of 2nd order	Ignition intervals
6-cylinder (continued)					
 V 90°, 3 throws	0	0	$\sqrt{3} \cdot F_1 \cdot a$ [2]	$\sqrt{6} \cdot F_2 \cdot a$	150°/90° 150°/90°
 Normal balance V 90°, 3 throws, 30° crank offset	0	0	$0.4483 \cdot F_1 \cdot a$	$(0.966 \pm 0.256) \cdot \sqrt{3} \cdot F_2 \cdot a$	120°/120°
 Opposed-cylinder (boxer), 6 throws	0	0	0	0	120°/120°
 V 60°, 6 throws	0	0	$3 \cdot F_1 \cdot a/2$	$3 \cdot F_2 \cdot a/2$	120°/120°
8-cylinder					
 V 90°, 4 throws in two planes	0	0	$\sqrt{10} \cdot F_1 \cdot a$ [2]	0	90°/90°
12-cylinder					
 V 60°, 6 throws	0	0	0	0	60°/60°

[1] Without counterweights, [2] Can be completely balanced by using counterweights.

Tribology and friction

The piston with piston rings and bearing surface is a self-contained and highly complex tribological system. This also applies to the piston pin and the crankshaft friction bearings. Detailed simulation formulations are necessary in view of the periodically changing application of force both in terms of amount and direction.

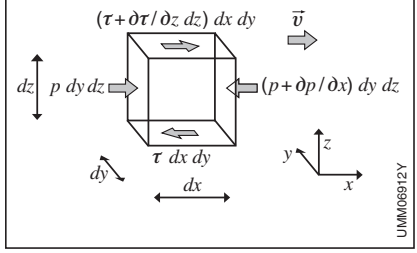
To be able to ensure wear resistance, it is necessary to build up a hydrodynamic lubricating film which is higher than the surface roughness of the two contact surfaces (see also Stribeck curve). The oil film either is already present in the lubricating gap (friction bearing with compressed-oil supply) or must be built up dynamically as an oil wedge (cambered piston ring on liner). However, the oil wedge always collapses when the relative speed of the contact surfaces assumes a value of zero. This is the case for piston rings at the piston reversal points (top and bottom dead centers), and for the piston pin between these positions. It is therefore also important to ensure that a sufficient oil-retaining volume can be maintained in the contact surfaces in order to retain by forces of adhesion in the short time of standstill as much oil as to avoid seizing or minimize increased wear. In the tribological system for the piston this is bearing-surface honing, and for the piston pin shaping of the piston-pin bearing.

Traced back to Newton is the formulation to the effect that the friction between two adjacent oil elements is virtually independent of the prevailing pressure and only proportional to the change in speed from one element to the next.

$$\tau = \frac{\eta\, dv}{dz}.$$

$\tau = F/A$ Shear stress,
F Shear force,
A Contact surface,
v Speed,
z Coordinate direction normal to speed direction v.

Figure 19: Pressure forces in a one-dimensional stress state



Sometimes the kinematic viscosity v is also used; this corresponds to the dynamic viscosity referred to density and is defined by

$$v = \frac{\eta}{\rho}\ [\mathrm{m^2/s}].$$

The dynamic viscosity η [N s/m^2 or Pa · s] decreases sharply with temperature. It is important to see to it that the oil temperature ensures a viscosity η which facilitates sufficient shear stress to build up a supporting oil film. It is also necessary to provide for a sufficiently high relative speed v on the part of the contact surfaces. It can be recognized from the formula that hydrodynamic shear stress collapses when the speed assumes a value of zero (reversal point of the piston, pivoting-angle end of the small connecting-rod eye). If can further be recognized that the lubricating film Δy must not become too large, as otherwise the shear stress will become too small and the lubricating film will break.

The equilibrium of forces for the simplified case of a one-dimensional stress state in the x direction (Figure 19) shows that $\partial p/\partial x = \partial\tau/\partial z$ must apply and thus $\partial p/\partial x = \eta\, \partial^2 v/\partial z^2$. It is thus demonstrated on the one hand that the pressure loss of a flow is dependent on the viscosity and the speed. In addition, the speed profile can thus be derived.

On account of the adhesion condition which states that the speed of the oil on the wall is identical to the wall speed, a speed distribution (linear for laminar flow) and a pressure distribution build up in the lubricating gap (see Friction bearings).

It is attempted with the aid of (thermo-) elastohydrodynamic simulation (EHD, elastohydrodynamics) to determine the lubricating-film thickness by solving the mathematical equations in the multi-dimensional case for mass balance and momentum equilibrium.

– Mass balance: supplied oil mass and oil quantity escaping through the bearing gap in equilibrium.
– Momentum equilibrium: equilibrium of forces from direct stress (pressure force from oil pressure), shear stress (viscosity force from oil viscosity) including apparent viscosity stresses on account of turbulent flow and the inertial force (acceleration component).

The equation system is established as a Reynolds differential equation or a Navier-Stokes equation to solve the flow problem. By using a finite-element simulation, the bearing deformations are added and the highly complex equations (because they are simultaneous) are numerically solved.

An adequate supply of oil must be guaranteed in the bearings, since generally the oil pressure in the bearings (> 100 bar) far exceeds the static oil pressure provided by the oil pump (< 10 bar). This is achieved by means of pressure-oil bores and special grooves in friction bearings or by means of initial forces at the piston rings.

Attempts have also been made in the meantime to simulate the case of mixed friction. The influence of surface fine geometry (consideration of roughness on microhydrodynamics by “flow tensors”) is taken into account in an extended Reynolds differential equation. By using contact-pressure models of rough surfaces, it is possible to determine, as well as the hydrodynamic contact-area ratios, the solid contact-area ratio and to take this into account accordingly. Intensive work is being conducted on models to determine the running-in behavior of contact surfaces. This will finally enable the wear behavior to be assessed.

Because the shear stresses and thus also the frictional forces are proportional to the speed, the share of friction of the piston assembly at the dead centers and that of the piston-pin bearing is small. Instead, the wear parameters are decisive here.

Sliding bearings

The bearings most commonly encountered in an internal-combustion engine are friction (sliding) bearings. Crankshaft main bearings, big-end bearings, piston-pin bearings, and camshaft bearings are generally oil-lubricated friction bearings (see Friction bearings).

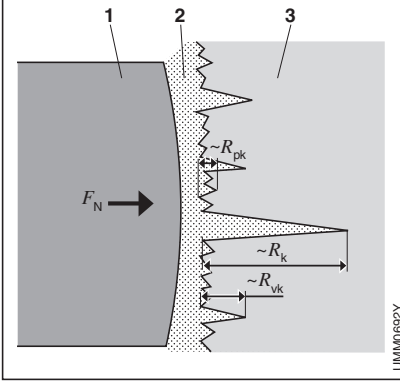
It is important to ensure that the seating pressure is precisely configured in both split bearings (crankshaft, connecting rod) and non-split bearings (piston pin, camshaft) so as to prevent them from jamming and rotating also at higher temperatures or under emergency-lubrication conditions. However, material overloading must also be reliably prevented.

The lubricating gap of the crankshaft main bearings changes in response to the constantly varying bearing forces (see also Crankshaft-assembly design). This generates a kind of pumping effect and the oil in the bearing is constantly replaced.

The swinging motion of the connecting rod effects an increased relative speed when the piston travels up the cylinder and a reduced relative speed when the piston travels down the cylinder in the lubricating gap of the big-end bearing or crank pin. This alone causes oil to be pumped into the oil gap. Furthermore, naturally periodic forces acting on the bearing cause the connecting rod to change position constantly and thereby assist the build-up of lubricating-film pressure.

The small-end bearing is merely subject to a swinging motion. A supply of oil is therefore particularly important, but difficult to achieve. In practical terms, mixed friction cannot be avoided. If the arrangement involves a floating pin bearing (bearing in the piston and connecting rod), the pin performs a slight rotary motion in the

Figure 20: Roughness values and oil-film theory (roughness definitions, see DIN 4760 [5])
1 Piston, 2 Oil film, 3 Liner.



piston bearing of just a few degrees – here too a good supply of oil to the bearings must be ensured.

Oil is generally supplied as splash oil, and partly also via a specific locating bore or via a pressure-oil supply from the crankshaft through grooves in the crankshaft-journal bearing and a vertical bore in the connecting rod.

Rolling pairs
Ball, roller or needle bearings are used in 2-stroke engines or sometimes in camshafts (see Roller bearings).

A roller follower or the roller of an injection-pump plunger-and-barrel assembly is a special case in point. The system of rotating roller and moving cam is a rolling-element pair with linear contact. A characteristic load variable is the Stribeck rolling-contact pressure:

$$K = \frac{F}{D_l \, l_{eff}},$$

F Force load,
$\varphi = D_l/D$ Osculation coefficient,
D_l Substitute diameter,
D Roller diameter,
l_{eff} Effective roller width.

The effective substitute diameter D_l is obtained from the instantaneous bend radius of the cam R_{cam} and the roller radius R_{roller}:

$$D_l = 2 \, \frac{R_{cam} \, R_{roller}}{R_{cam} + R_{roller}}.$$

It takes into account the osculation of the two rolling bodies.

The Hertzian stress p_{Hertz} is also relevant:

$$p_{Hertz} = \sqrt{\frac{KE}{2.86}}.$$

with modulus of elasticity E:

$$E = 2 \, \frac{E_{cam} \, E_{roller}}{E_{cam} + E_{roller}}.$$

By slightly angling the roller in relation to the cam or by offsetting the stroke function, it is possible to ensure that an oil wedge can be built up and the roller rotates continuously. Pitting is a typical indicator of overloading in rolling pairs that are subject to high loads. It results from molecular deformations of the metal and possible embedding of oil, which gives rise to extremely high local pressures and plastic deformations or material erosion. Harder materials and decreasing the load reduce the tendency to pitting.

Piston ring and bearing surface
The piston rings together with the bearing surface constitute a complex tribological system. On the one hand, the pressure force of the piston ring is very much dependent on the cylinder pressure and thus on the crank angle and the power stroke. On the other hand, the piston rings make axial secondary movements in their grooves which also have a significant impact on the lubricating-film thickness.

The oil is generally applied as splash oil to the bearing surface (crankshaft churning or splash oil for piston cooling and pin lubrication) and allowed back through the oil control ring as a thin film. The oil-retaining volume of the liner is dependent on the honing. This includes the surface roughness (furrow depth) and the honing pattern (cross-furrow angle).

Honing the bearing surface serves two functions. Firstly, it is intended to ensure with the aid of honing furrows of sufficient depth that oil can be retained by adhesion forces for an adequate length of time (Figure 20). Secondly, surface roughness must be so low as to prevent contact with the piston ring. Various roughness

values are used to characterize honing; these values are standardized on a generic level in DIN 4760 [5] and in detail in DIN EN ISO 13565 [6]. For initial wear and initial friction the R_{pk} value ("reduced peak height", see DIN 4760 and DIN EN ISO 11562 [7] or DIN EN ISO 13565 for an explanation) is decisive – but the value is swiftly and drastically minimized after run-in (e.g. from a few micrometers to 0.2 to 0.8 µm). The R_{vk} value (1.0 to 3.5 µm) is the decisive parameter for the possible oil-retaining volume after run-in. It is to be noted that the honing usually takes the form of a cross-shaped furrow system. In this form, the furrows constitute a communicating system of oilways and the oil pressure is partly forced out of the gap, but this also permits a uniform distribution of the oil film over the surface. Flat honing angles (referring to the angle formed by the cross-shaped furrows) reduce the amount of oil forced out, but there is a tendency to seizing on account of the unsatisfactory axial oil distribution. Steep honing angles result in increased oil consumption. Honing angles between 30° and 90° are customary.

Mixed friction occurs at any rate at the piston reversal points. The least that is required in this area is micro-pockets of oil, which can be produced for example by laser honing or by bringing out amounts of alloy during the honing procedure.

Usually, such high temperatures are generated during the combustion stroke of a piston engine that much of the bearing-surface oil vaporizes or in diesel engines even burns off. This vaporized oil constitutes the majority of the oil consumption by modern-day internal-combustion engines.

The special shape of the piston rings is designed to facilitate the build-up of an oil wedge. A critical factor is, on the one hand, the piston speed, which must not be allowed to become too high, as otherwise the shear forces can no longer be absorbed by the oil and the oil film will break. On the other hand, the reversal points, particularly at top dead center, must be classed as critical with regard to wear.

Empirical values and data for calculation

Characteristics

Fuel consumption

Figure 1 shows in a diagram nine different specific consumptions plotted against the effective mean pressure p_{me}. Since p_{me} has a proportional effect on the equation of energy ($E_E = i\,p_{me}\,V_H$; $i = 1$ for two-stroke engines, i = 0.5 for four-stroke engines), the absolute consumption rises with increasing p_{me}. This situation is illustrated for example by straight line 1.

The specific consumptions rise from straight lines 1 through 9. In other words, for the first process 200 g of fuel are needed to obtain 1 kWh of energy. For 9, 500 g of fuel already are expended for 1 kWh.

The y-axis is normalized to the working cycle and the displacement (swept volume). In other words, the energy E_E is directly dependent on p_{me}. When p_{me} increases, so too the energy increases. More fuel is needed when the energy increases. If a certain p_{me} is taken – e.g. 15 bar, which corresponds to a defined energy – significantly more fuel is required for straight line 5 than for straight line 1 (roughly 120 mg as against 80 mg) to generate the same amount of mechanical energy. This is also discernible from the gradient of the straight line.

Characteristic values of gasoline and diesel engines

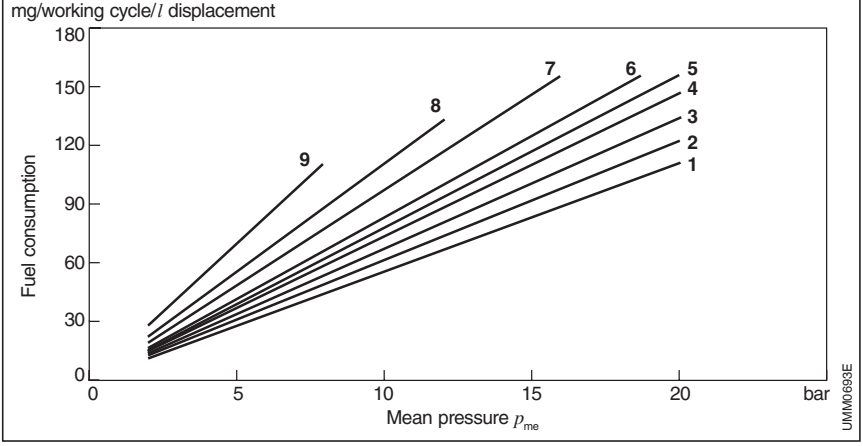
Table 1 shows characteristic values in a comparison of gasoline and diesel engines. What is noticeable at the outset is the large differences in rpm between gasoline and diesel engines. The compression ratios too are – by the nature of the process – much higher for diesel engines than for gasoline engines. Due to the higher compression pressure, greater mean pressures p_{me} are possible. Since diesel engines attain lower speeds (rpm), they have a lower power output per liter (power output per displacement). Due to the lower power output per liter and the resulting higher weight for durability reasons, diesel engines have the higher weight-to-power ratio (vehicle mass per power output). Due to the lower compression ratio, gasoline engines have an unfavorable specific consumption.

Torque position

The torque position denotes the position on the engine-speed curve (Figure 1) at which maximum torque ($n_{Md,max}$) is developed, relative to the rated-power speed

Figure 1: Fuel consumption

1 200 g/kWh, 2 220 g/kWh, 3 240 g/kWh,
4 260 g/kWh 5 280 g/kWh, 6 300 g/kWh,
7 350 g/kWh, 8 400 g/kWh, 9 500 g/kWh.



(n_{nom}) in percent ($n_{Md,max}/n_{nom}$). For typical values, see Table 2.

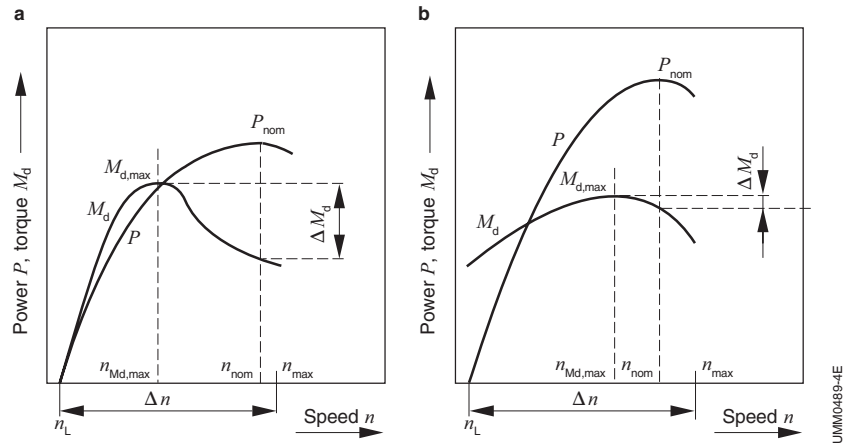
Useful speed range
The useful speed range Δn shows the range in which an engine is operated. For example, a gasoline engine with a maximum speed n_{max} of 7000 rpm and an idle speed of 800 rpm has a useful speed range of 6200 rpm. The figures for useful speed range in Table 2 are given in rpm x 1000.

Table 1: Comparative data

Engine type and application			Engine speed n_{nom} rpm	Compression ratio ε	Max. mean pressure p_{me} bar	Power output per liter kW/l	Weight-to-power ratio kg/kW	Specific fuel consumption g/kWh	Torque increase %
Gasoline engines	Motor-cycles	4-stroke	5000…13000	9…12	9…13	50…150	2.5…0.5	230…280	10…15
	Pass. cars	NAE	5000…8000	9…13	11…14	40…80	2.0…0.8	220…270	15…20
		SCE/ IC	5000…7500	9…12	15…22	60…110	1.5…0.5	220…250	20…40
Diesel engines	Pass. cars, light comm. veh.	SCE/ IC	3500…4500	18…22	12…20	35…55	3.0…1.3	200…220	20…40
	Comm. veh.	SCE/ IC	1800…2600	15…18	18…24	25…40	4.0…2.5	180…210	20…40

NAE Naturally aspirated engine; SCE Supercharged engine; IC Intercooling (charge-air cooling)

Figure 2: Typical power and torque curves
a) Diesel engine with exhaust-gas turbocharging,
b) Gasoline engine (naturally aspirated engine).
n_L Idle speed, $n_{Md,max}$ Engine speed at $M_{d,max}$, n_{nom} Nominal speed, n_{max} Maximum engine speed, Δn Useful speed range, P_{nom} Rated power, $M_{d,max}$ Maximum torque.



Torque rise
The torque rise ΔM_d denotes the difference between the maximum torque $M_{d,max}$ and the torque at the rated-power point. The values in Table 3 are given in percent.

Torque increase
The torque increase is the gradient of the torque curve between maximum torque and the torque at rated power.

Engine output and atmospheric conditions
The torque and thus the power output of an internal-combustion engine are essentially determined by the heat content of the cylinder charge. The flow rate of air (or, more precisely, of oxygen) in the cylinder charge provides a direct indication of heat content. The change which the engine will display at full power can be calculated as a function of variations in the condition of the ambient air (temperature, barometric pressure, humidity), provided that engine speed, air/fuel (A/F) ratio, volumetric efficiency, combustion efficiency, and total engine power loss remain constant. The A/F mixture responds to lower atmospheric density by becoming richer. The volumetric efficiency (ratio of pressure in cylinder at BDC to pressure in ambient air) only remains constant for all atmospheric conditions at maximum throttle-valve aperture (WOT). Combustion efficiency drops in cold thin air as vaporization rate, turbulence, and combustion speed all fall. Engine power loss (friction losses + gas-exchange work + boost power drain) reduces the indicated power.

Table 2: Typical values for useful speed range and torque position

Engine type		Useful speed range Δn_N (×1000 rpm)	Torque position [%]
Diesel engine	Pass. cars	3.5…5	15…40
	Comm. veh.	1.8…3.2	10…60
Gasoline engine		4…7	25…35

Table 3: Typical values for torque rise

Engine type		Torque rise ΔM_d [%]
Diesel engine, pass. cars	SCE and IC	25…35
Diesel engine Comm. veh.	SCE and IC	25…40
Gasoline engine	Nat. asp. engine	25…30
	SCE and IC	30…35
SCE Supercharged engine; IC Intercooling (charge-air cooling)		

Effect of atmospheric conditions
The quantity of air which an engine draws in, or is inputted to the engine by supercharging, depends upon ambient-air density; colder, heavier, denser air increases engine output. As a rule of thumb: Engine power drops by approximately 1 % for each 100 m rise in elevation. Depending upon engine design, the cold intake air is normally heated to some degree while traversing the intake passages, thereby reducing its density and thus the engine's ultimate output. Humid air contains less oxygen than dry air and therefore produces lower engine power outputs. The decrease is generally modest to the point of insignificance. The warm humidity of air in tropical regions can however result in a noticeable engine power loss.

Definitions of power
The effective power is the engine's power as measured at the crankshaft or ancillary mechanism (such as the transmission) at the specified engine speed. When measurements are made downstream from the transmission, the transmission losses must be factored into the equation.

The rated power is the maximum effective power of the engine at full throttle. The net power corresponds to the effective power.

Conversion formulas are used to convert the results of dynamometer testing to reflect standard conditions, thereby negating the influences of such factors as time of day and year while simultaneously allowing the various manufacturers to provide mutually comparable data. The procedure converts air density – and thus the effective volume of air in the engine – to defined "standard conditions" for air mass.

Table 4: Power-output correction factors

Gasoline engines, naturally aspirated and supercharged		
	EEC 80/1269 [9], JIS D 1001 [10], SAE J 1349 [11]	DIN 70020 [12]
Correction factor α_a	$\alpha_a = A^{1.2} \cdot B^{0.5}$ $A = 99/p_{PT}$ $B = T_p/298$	$\alpha_a = A \cdot B^{0.5}$ $A = 101.3/p_{PF}$ $B = T/293$
Corrected power: $P_0 = \alpha_a P$ (kW) (P measured power)		
Diesel engines		
Atmospheric engine correction factor f_a	$f_a = A^{0.7} \cdot B^{1.5}$ ($A = 99/p_{PT}$; $B = T_p/298$) (turbocharged engines with/without intercooling)	As α_a for gasoline engines
Engine correction factor f_m	$40 \le q/r \le 65$: $f_m = 0.036\,(q/r) - 1.14$ $q/r < 40$: $f_m = 0.3$ $q/r > 65$: $f_m = 1.2$	$f_m = 1$
$r = p_L/p_E$ Charge-air pressure response, with p_L absolute charge-air pressure, p_E Absolute pressure upstr. of compressor, q Spec. fuel consumption (SAE J 1349, [8]), p_{PT} Barometer pressure dry, p_{PF} Barometer pressure absolute, T_P Test-dynamometer temperature. 4-stroke engines: $q = 120{,}000\ F/DN$; 2-stroke engines: $q = 60{,}000\ F/DN$, with F Fuel flow [mg/s], D Displacement [l]; N Engine speed [rpm].		
Corrected power: $P_0 = P f_a^{f_m}$ (kW) (P measured power)		

The comparison data in Table 4 show the most important factors for these power-output corrections.

Displacement
The displacement – also referred to as the swept volume – is characterized by the total vertical stroke of the piston multiplied by the cylinder diameter. The displacement of an individual cylinder is characterized as V_h. The displacement V_H of an engine corresponds to the sum total of all the displacements V_h of its cylinders.

Compression volume
The compression volume V_c corresponds to the compression space in the cylinder when the piston is at TDC.

Combustion chamber
The combustion chamber is obtained from the sum total of displacement and compression volume.

Compression ratio
The compression ratio describes the ratio of the combustion-chamber volume when the piston is at BDC to the volume when the piston is at TDC.

Piston stroke
The piston stroke is a function of the crank angle. It is not uniform over one revolution and varies. Thus, the piston moves from TDC towards BCD up to 90° CA after TDC much faster downwards than from 90° to BDC. The mean piston speed is the average speed of the piston stroke over one complete revolution.

Efficiency
The efficiency describes the ratio of the mechanically supplied power to the energy supplied chemically by the fuel.

Calculations
Table 6 shows calculation formulas for some quantities in the engine. The quantity equations are universally valid, the numerical values in the given unit are inserted in the numerical-value equations. These equations tend to be applied in practice.

The frequently used quantities and their units are set out in Table 5.

Table 5: Quantities and units

Quantity		Unit
a_K	Piston acceleration	m/s²
B	Fuel consumption	kg/h; dm³/h
b_e	Spec. fuel consumption	g/kWh
d	Cylinder diameter	mm
d_v	Valve diameter	mm
F	Force	N
F_G	Gas force in the cylinder	N
F_N	Piston side thrust	N
F_o	Oscillating inertial force	N
F_r	Rotating inertial force	N
F_s	Connecting-rod force	N
F_T	Tangential force	N
M	Torque	Nm
M_o	Oscillating moments	Nm
M_r	Rotating moments	Nm
M_d	Engine torque	Nm
m_p	Weight-to-power ratio	kg/kW
n	Engine speed	rpm
n_p	Fuel-injection pump speed	rpm
P	Power	kW
P_{eff}	Effective power	kW
P_H	Power output per liter	kW/dm³
p	Pressure	bar
p_c	Final compression pressure	bar
p_e	Mean piston pressure (mean pressure, mean working pressure)	bar
p_L	Charge-air pressure	bar
p_{max}	Peak injection pressure in the cylinder	bar
r	Crankshaft radius	mm
s_d	Injection cross-section of the nozzle	mm²
S, s	Stroke, general	mm
s	Piston stroke	mm
s_f	Induction stroke of a cylinder (2-stroke)	mm
s_F	Induction stroke, 2-stroke engine	mm
S_k	Piston clearance from TDC	mm
S_s	Slot height, 2-stroke engine	mm
T	Temperature	°C, K
T_c	Final compression temperature	K
T_L	Charge-air temperature	K
T_{max}	Peak temperature in combustion chamber	K
t	Time	s
V	Volume	m³
V_c	Compression volume of a cylinder	dm³
V_E	Injected fuel quantity per pump stroke	mm³
V_f	Charge volume of a cylinder (2-stroke)	dm³
V_F	Charge volume of a 2-stroke engine	dm³
V_h	Displacement of a cylinder	dm³
V_H	Displacement of the engine	dm³
v	Velocity	m/s
v_d	Mean velocity of injection spray	m/s
v_g	Gas velocity	m/s
v_m	Mean piston speed	m/s
v_{max}	Max. piston speed	m/s
z	Number of cylinders	–
β	Pivoting angle of connecting rod	°
ε	Compression ratio	–
η	Efficiency	–
η_e	Net efficiency	–
η_{th}	Thermal efficiency	
ν, n	Polytropic exponent of real gases	–
ρ	Density	kg/m³
φ, α	Crank angle (φ_o = top dead center)	°
ω	Angular velocity	rad/s
λ	= r/l Stroke/connecting-rod ratio	–
λ	Excess-air factor	–
κ	= c_p/c_v Adiabatic exponent of ideal gases	–

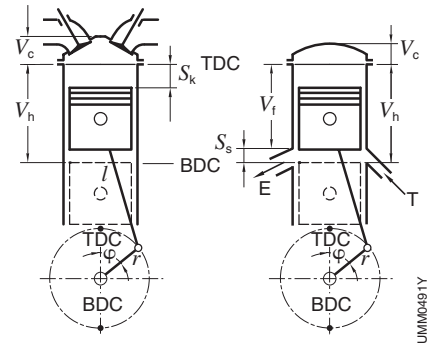
Superscripts and subscripts

0, 1, 2, 3, 4, 5	Cycle values/main values
o	Oscillating
r	Rotating
1st, 2nd	1st, 2nd order
A	Constant

Conversion of units

1 g/hp·h	= 1.36 g/kWh
1 g/kWh	= 0.735 g/hp·h
1 kpm	= 9.81 Nm ≈ 10 Nm
1 Nm	= 0.102 kpm ≈ 0.1 kpm
1 hp	= 0.735 kW
1 kW	= 1.36 hp
1 at	= 0.981 bar ≈ 1 bar
1 bar	= 1.02 at ≈ 1 at

Table 6: Calculation equations

Mathematical relationship between quantities	*Numerical relationship between quantities*
Displacement Displacement of a cylinder $V_h = \frac{\pi d^2 s}{4}$ (4-stroke); $V_f = \frac{\pi d^2 s_f}{4}$ (2-stroke) Displacement of the engine $V_H = V_h z$ (4-stroke); $V_F = V_f z$ (2-stroke)	$V_h = 0.785 \cdot 10^{-6} d^2 s$ V_h in dm^3, d in mm, s in mm $V_H = 0.785 \cdot 10^{-6} d^2 s z$ V_h in dm^3, d in mm, s in mm
Compression Compression ratio $\varepsilon = \frac{V_h + V_c}{V_c}$ Final compression pressure $p_c = p_o \varepsilon^v$ Final compression temperature $T_c = T_o \varepsilon^{v-1}$	 4-stroke engine 2-stroke engine
Piston stroke Piston clearance from top dead center $S_k = r\left[1 + \frac{l}{r} - \cos\varphi - \sqrt{\left(\frac{l}{r}\right)^2 - \sin^2\varphi}\right]$ Crank angle $\varphi = 2\pi n t$ (φ in rad) Piston speed (approximation) $v \approx 2\pi n r\left(\sin\varphi + \frac{r}{2l}\sin 2\varphi\right)$ Mean piston speed $v_m = 2ns$ Maximum piston speed (approximate, if connecting rod is on a tangent with the big-end trajectory; $a_k = 0$	$\varphi = 6 n t$ φ in °, n in rpm, t in s $v \approx \frac{ns}{19100}\left(\sin\varphi + \frac{r}{2l}\sin 2\varphi\right)$ v in m/s, n in rpm, l, r and s in mm $v_m = \frac{ns}{30000}$ v_m in m/s, n in rpm, s in mm
Piston acceleration (approximation) $a_k \approx 2\pi^2 n^2 s\left(\cos\varphi + \frac{r}{l}\cos 2\varphi\right)$	$a_k \approx \frac{n^2 s}{182400}\left(\cos\varphi + \frac{r}{l}\cos 2\varphi\right)$ a_k in m/s^2, n in rpm, l, r and s in mm

l/r	3.5	4	4.5
v_{max}	1.63 v_m	1.62 v_m	1.61 v_m

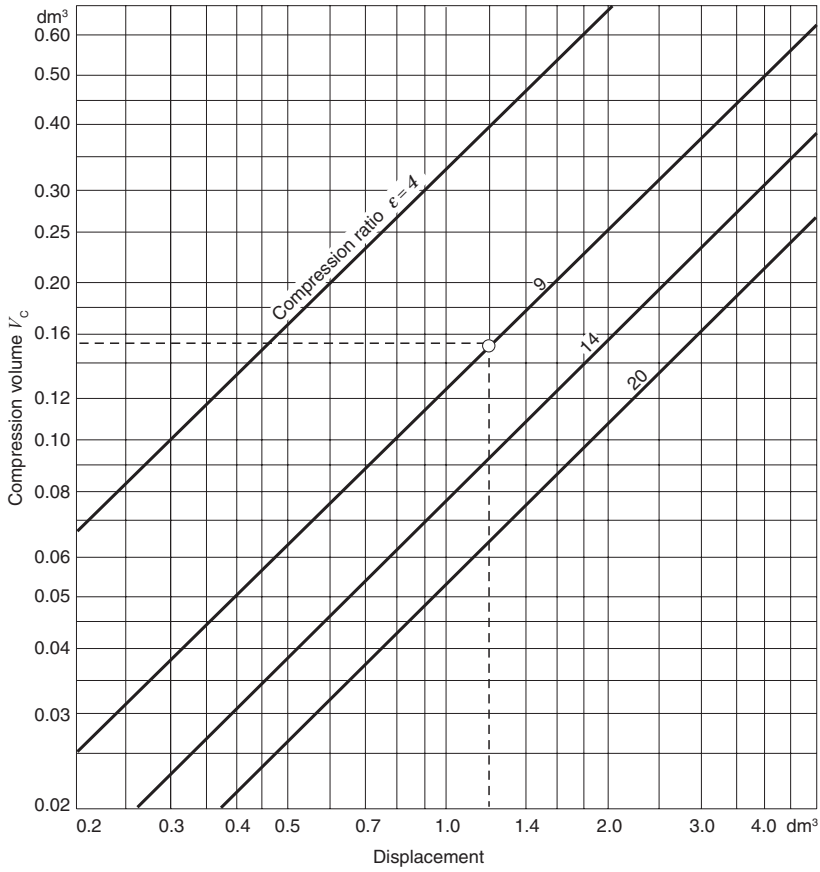
Table 6: Calculation equations (continued)

Mathematical relationship between quantities	*Numerical relationship between quantities*
Gas velocity Mean gas velocity in the valve section $v_g = \frac{d^2}{d_v^2} v_m$	$v_g = \frac{d^2}{d_v^2} \cdot \frac{n s}{30\,000}$ v_g in m/s, d, d_v, and s in mm, n in rpm
Engine power $P = M \omega = 2 \pi M n$	$P = \frac{M n}{9549}$ P in kW, M in N m (= W s),
$P_{eff} = V_H p_e \frac{n}{K}$ K = 1 for 2-stroke engine K = 2 for 4-stroke engine	$P_{eff} = \frac{V_H p_e n}{K \cdot 600} = \frac{M_d n}{9549}$ P_{eff} in kW, p_e in bar, n in rpm M_d in N m
Power per unit displacement (power output per liter) $P_H = \frac{P_{eff}}{V_H}$	$P = \frac{M n}{716.2}$ P in hp, M in kp · m, n in rpm
Weight-to-power ratio $m_p = \frac{m}{P_{eff}}$	

Table 6: Calculation equations (continued)

Mathematical relationship between quantities		*Numerical relationship between quantities*	
Mean piston pressure (mean pressure, mean working pressure)			
4-stroke engine $p = \frac{2P}{V_H n}$	2-stroke engine $p = \frac{2P}{V_H n}$	4-stroke engine $p = 1{,}200 \frac{P}{V_H n}$	2-stroke engine $p = 600 \frac{P}{V_H n}$
		p in bar, P in kW, V_H in dm³, n in rpm	
		$p = 882 \frac{P}{V_H n}$	$p = 441 \frac{P}{V_H n}$
		p in bar, P in hp, V_H in dm³, n in rpm	
$p = \frac{4\pi M}{V_H}$	$p = \frac{2\pi M}{V_H}$	$p = 0.1257 \frac{M}{V_H}$	$p = 0.0628 \frac{M}{V_H}$
		p in bar, M in N m, V_H in dm³	
Engine torque			
$M_d = \frac{V_H P_e}{4\pi}$	$M_d = \frac{V_H P_e}{4\pi}$	$M_d = \frac{V_H P_e}{0.12566}$	$M_d = \frac{V_H P_e}{0.06284}$
		M_d in N m, V_H in dm³, p_e in bar	
		$M_d = 9{,}549 \frac{P_{eff}}{n}$	
		M_d in N m, P_{eff} in kW, n in rpm	
Fuel consumption			
B = Measured values in kg/h $b_e = \frac{B}{P_{eff}}$ $b_e = \frac{1}{H_u \eta_e}$		B in dm³/h or kg/h V_B = Measured volume on test dynamometer t_B = Elapsed time for measured volume consumption $b_e = 3{,}600 \frac{V_b \rho_B}{t_b P_{eff}}$ ρ_B = Fuel density in g/cm³, t_B in s, V_B in cm³, P_{eff} in kW.	
Efficiency			
$\eta_{th} = 1 - \varepsilon^{1-k}$ $\eta_e = \frac{P_{eff}}{B H_u)}$		$\eta_e = \frac{x}{b_e}$ $x = 82$ for $H_u = 44$ $x = 86$ for $H_u = 42$ $x = 90$ for $H_u = 40$ $x = 120$ for $H_u = 30$ H_u Specific calorific value in MJ/kg b_e Specific fuel consumption in g/kWh	

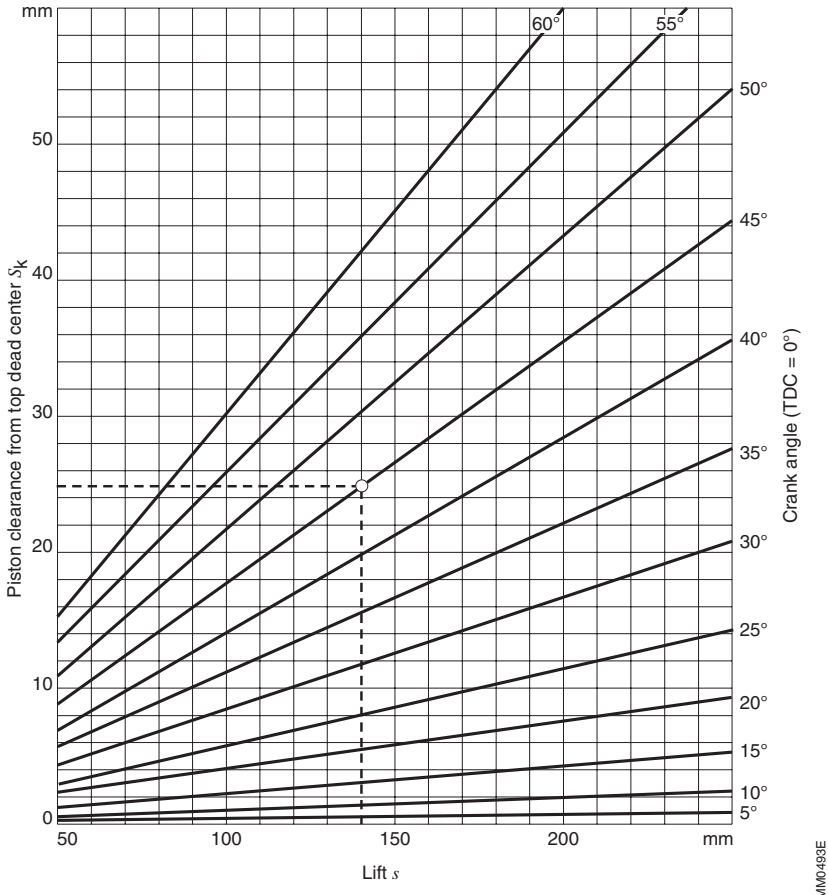
Figure 3: Displacement and compression space
The diagram applies to the displacement V_h and the compression space V_c of the individual cylinder and to the total displacement V_H and the total compression space V_C.



Displacement and compression space
In Figure 3 the displacement is plotted on the x-axis and the compression space is plotted on the y-axis. The straight lines show lines with a constant compression ratio. As the displacement increases, so too the compression space increases. If the compression ratio is raised, the compression space must become smaller for the same displacement.

An engine with a displacement of 1.2 dm³ and a compression ratio ε = 9 has a compression space of approximately 0.15 dm³.

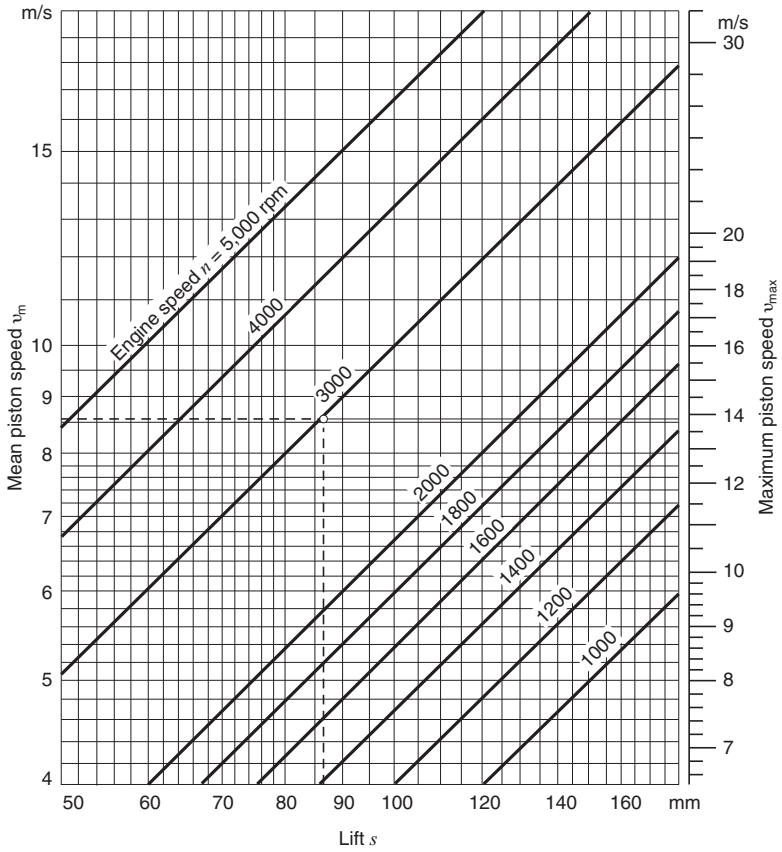
Figure 4: Piston clearance from top dead center
Conversion of degrees of crankshaft angle to mm piston travel.



Determining the piston clearance
Example
For a stroke of 140 mm the piston clearance from top dead center at 45 °CA is 25 mm (Figure 4).

The diagram is based on a crank ratio $l/r = 4$ (l connecting-rod length, r on half of the stroke length). However, it applies with good approximation (error less than 2 %) to all crank ratios l/r between 3.5 and 4.5.

Figure 5: Piston speed



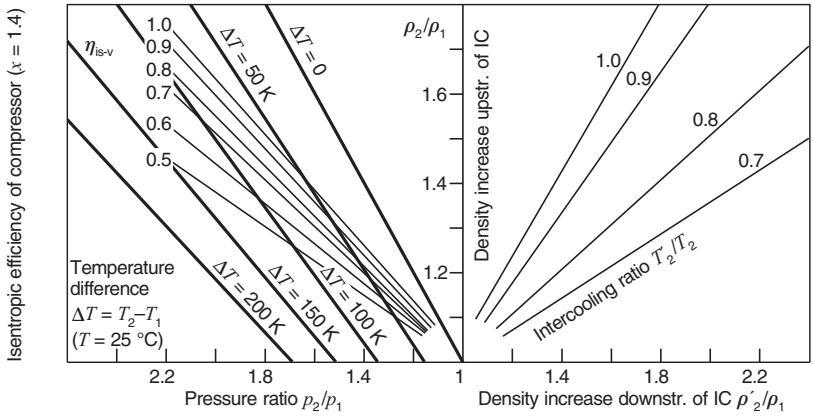
Determining the piston speed
The x-axis in Figure 5 shows the stroke in millimeters. The left y-axis shows the mean and the right y-axis the maximum piston speed. In addition, lines of constant speed are marked. It can clearly be seen that both piston speeds increase as the stroke increases. The same applies to increasing speeds.

Example
For the stroke s = 86 mm and at an engine speed of n = 3000 rpm the mean piston speed v_m = 8.7 m/s and the maximum piston speed
v_{max} = 13.8 m/s.

The diagram is based on $v_{max} = 1.62\, v_m$.

Figure 6: Density increase of combustion air in cylinder on turbo/supercharging

Increase in density on supercharging as a function of the pressure ratio in the compressor, of the compressor efficiency, and of the intercooling rate for intercooling (IC).

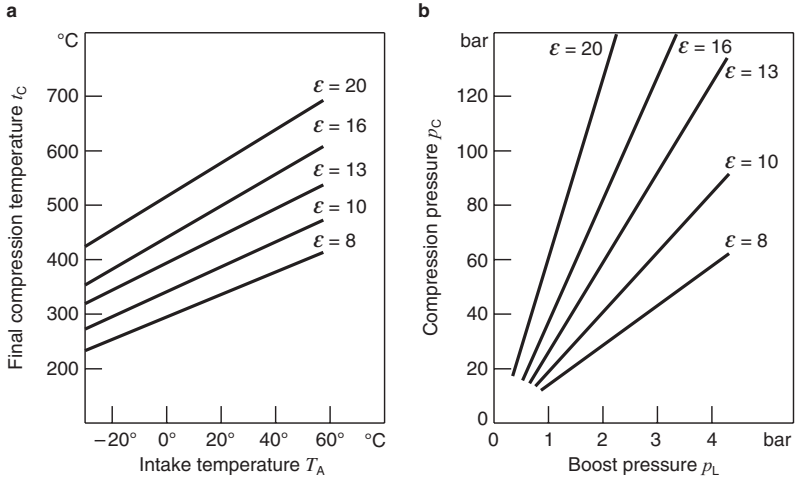


$p_2/p_1 = \pi_c$ = Pressure ratio during precompression,
ρ_2/ρ_1 = Increase in density, ρ_1 = Density upstream of compressor, ρ_2 = Density downstream of compressor in kg/m³,
T_2'/T_2 = Intercooling rate, T_2 = Temperature before IC, T_2' = Temperature after IC in K,
η_{is-v} = Isentropic compressor efficiency.

UMM0495-3E

Figure 7: Final compression pressure and final compression temperature

a) Final compression temperature as a function of compression ratio and intake temperature,
b) Final compression pressure as a function of compression ratio and charge-air pressure.



$t_c = T_c - 273{,}15$ K, $T_c = T_A\, \varepsilon^{n-1}$, n = 1,35

$p_c = p_L\, \varepsilon^n$, n = 1.35

UMM0496E

Determining the density increase of the combustion air

Figure 6 shows that as the pressure ratio increases, depending on how strongly the air is heated and cooled in the intercooler, the density of the gas increases. The hotter the gas after the intercooler, the smaller the density increase.

Determining the final compression pressure

Figure 7a shows on the x-axis the intake temperature and on the y-axis the final compression temperature. In addition, lines of a constant compression ratio are marked. It can be seen that the final compression temperature increases as the intake temperature increases. The same applies to increasing compression ratios, where here the straight-line gradient increases with a higher compression ratio. This graphic shows why lowering the compression ratio can reduce the knock tendency.

Figure 7b shows the final compression pressure as a function of compression ratio and charge-air pressure. If both increase, the final compression pressure also increases. However, a significant effect on the final compression pressure can be seen by increasing the compression ratio.

References

[1] H. Hütten: Motoren – Technik, Praxis, Geschichte. Motorbuch-Verlag, Stuttgart, 1997.
[2] DIN 73021: Designation of the rotational direction, the cylinders and the ignition circuits of motorcar engines.
[3] ISO 1204: Reciprocating internal combustion engines – Designation of the direction of rotation and of cylinders and valves in cylinder heads, and definition of right-hand and left-hand in-line engines and locations on an engine.
[4] I. Bronstein; K. Semendjajew: Taschenbuch der Mathematik. 7th Edition, Verlag Harri Deutsch, 2008.
[5] DIN 4760: Form deviations; Concepts, Classification system.
[6] DIN EN ISO 13565: Geometrical Product Specifications (GPS) – Surface texture: Profile method – Surfaces having stratified functional properties.
[7] DIN EN ISO 11562: Geometrical Product Specifications (GPS) – Surface texture: Profile method – Metrological characteristics of phase correct filters.
[8] SAE J 1349: Engine Power Test Code Spark Ignition and Compression ignition As Installed Net Power Rating.
[9] Council Directive 80/1269/EEC of 16 December 1980 on the approximation of the laws of the Member States relating to the engine power of motor vehicles.
[10] JIS D 1001: Road vehicles – Engine power test code.
[11] SAE J 1349: Engine Power Test Code Spark Ignition and Compression ignition As Installed Net Power Rating.
[12] DIN 70020: Road vehicles – Automotive engineering.

Exhaust-gas emissions

Historic measures

Endeavors at controlling air pollution are an absolutely integral element of environmental protection. It successes since the mid-20th century are thanks in particular to the economic analysis of the generation of pollutants and the derived further development of the combustion process, the consistent improvement of fuel quality, and the technical development of exhaust-gas treatment.

Smog and acid rain are still today common catchwords that have their origin in the second half of the 20th century. On region hugely affected by smog in that period was the US state of California. In order to reduce air pollution in areas of high population density such as Los Angeles, California had already enacted the first laws in the early 1960s (see [1]). Other states followed California's example. In 1970 the European Community for its part passed its first emission regulations. These legislative measures induced automobile manufacturers to fit their motor vehicles with exhaust-gas treatment systems in the markets affected. This has resulted in today's global market for exhaust-gas treatment systems.

Constituents of engine exhaust gas

Major constituents

Various combustion products are produced in internal-combustion engines during the combustion of fuel. Exhaust gas comprises primarily the non-toxic major constituents:

– nitrogen N_2 (constituent of the intake air),
– water vapor H_2O,
– carbon dioxide CO_2,
– oxygen O_2 (in diesel engines and gasoline engines with lean-burn operation).

Carbon dioxide is present in the air as a natural constituent and is not categorized as a pollutant in respect of exhaust-gas emissions for motor vehicles. However, it is considered to be one of the causes of the greenhouse effect and the associated global climate change. Since industrialization, the CO_2 content in the atmosphere has risen by roughly 40 % to today's figure of 400 ppm.

The amount of released carbon dioxide is directly proportional to fuel consumption. Measures taken to reduce fuel consumption are therefore becoming increasingly significant.

Pollutants

The above-mentioned constituents are the components of exhaust gas that would be produced in a complete, ideal combustion of pure fuel – i.e. in the complete combustion of fuel with oxygen without unwanted secondary reactions. However, minor constituents are also produced in real vehicle operation.

The quantity of minor constituents produced during combustion is very much dependent on the engine operating status. For the gasoline engine, the quantity of untreated exhaust gas (exhaust gas after combustion and before exhaust-gas treatment) amounts to approximately 1 % of the total amount of the exhaust gas for an engine at normal operating temperature and with a stoichiometric mixture composition (where $\lambda = 1$). The compo-

sition of diesel exhaust gas is very much dependent on the excess air.

The essential pollutants produced directly during the in-engine combustion of fuels are described in the following.

Hydrocarbons
Fuels that contain hydrocarbons never combust absolutely completely in engines. The incompletely combusted hydrocarbons emitted from an engine and their concentrations are determined in particular by the combustion and engine applications with injection strategy, injected fuel quantity, injection pressure, point of injection, etc., and by the air-fuel ratio λ, by the combustion temperature, and by further engine influencing factors.

Aliphatic hydrocarbons (alkanes, alkenes, alkines as well as their cyclic derivatives) are very nearly odorless. Cyclic aromatic hydrocarbons (e.g. benzene, toluene, polycyclic hydrocarbons) exude an odor. Partially oxidized hydrocarbons (e.g. aldehydes, ketones) have an unpleasant odor and form derived products when exposed to sunlight.

Carbon monoxide
Carbon monoxide is formed during the incomplete combustion of fuel due to an insufficient oxygen supply. This can be caused by locally rich areas as a result of insufficient mixture preparation (e.g. unvaporized fuel droplets or fuel film on the wall of the combustion chamber).

Carbon monoxide is a colorless, odorless, and tasteless gas. It is toxic since it binds to the hemoglobin in blood and thereby inhibits oxygen transport.

Particulates
Unlike gaseous exhaust-gas constituents, particulates are not a clearly defined chemical species. All solid or liquid combustion residues which can accumulate in a defined sampling process on a sample filter are called particulates. These are, aside from soot particles – consisting of a solid soot core and adherent incompletely combusted hydrocarbons resulting from fuel and lube oil – droplets of hydrocarbons, water and sulfurous acids (sulfates) as well as ash particles and abraded metal.

Sulfur oxides
The term sulfur oxides (SO_x) denotes a mixture of sulfur dioxide (SO_2) and sulfur trioxide (SO_3). Sulfur oxides are almost inevitably formed during the combustion of fuel containing sulfur. The sulfur content in the fuel must always be taken into consideration in the choice and design of exhaust-gas treatment systems in order to guarantee their long-term effectiveness.

Nitrogen oxides (nitric oxides)
The nitrogen content in fuels is generally low and plays a subordinate role in the formation of nitrogen oxides. During the combustion of fossil fuels nitrogen oxides are produced from reactions of intermediately formed hydrocarbon radicals and oxygen (O_2) with atmospheric nitrogen (N_2). Phenomenologically these reactions have hitherto not been explained in every detail despite broad empirical knowledge. The layout of the combustion chamber and the combustion application have a large influence, e.g. guidance of the flame front and its local temperatures.

The produced mixture of strongly oxidizing nitrogen oxides is termed “nitrous gases” and comprises nitrogen monoxide (NO), nitrogen dioxide (NO_2) and various dimers (e.g. N_2O_3, N_2O_4) which are in thermodynamic equilibrium. Laughing gas (N_2O) is not strongly oxidizing and is not classified as a nitrous gas. It is a strong greenhouse gas. When it comes to the frequently used generic term NO_x, it is always necessary to define whether laughing gas is included or not. Both definitions are common.

Basic principles of catalytic exhaust-gas treatment

It seems advantageous to convert the pollutants formed in the engine with each other or with another substance contained in the exhaust gas into less harmful substances. The following challenges must be surmounted for this purpose:

– In the fewest engine operating states the exhaust gas contains equivalent proportions of the substances to be able to convert these with each other.
– The exhaust-gas temperature is too low at many engine operating points to be able to get over the activation barrier of the desired reactions.

The strategy of catalytic exhaust-gas treatment starts at this point.

Catalytic exhaust-gas treatment
A catalyst is scientifically defined as a substance in the presence of which a certain reaction or a reaction sequence takes place more quickly under kinetically limited reaction conditions and which itself emerges unchanged. In other words, the catalyst reduces the activation barrier of the reaction catalyzed by it. However, the components of exhaust-gas treatment systems designated in the technical sense as a catalyst furthermore perform further functions (see Exhaust-gas treatment, Catalytic converter).

Catalyst components
Catalysts for exhaust-gas treatment consist of complex-structured, highly porous ceramics which are in the main applied as coatings (catalyst coating) to honeycombed carrier structures. A distinction is made between the following catalyst components.

Catalytically active components
Catalytically active components (e.g. noble metals) increase the speed of the desired reactions under suitable conditions.

Storage components
Storage components (e.g. zeolites) absorb certain substances from the exhaust gas and store them until suitable conditions exist to convert them.

Carrier components
Carrier components (e.g. Al_2O_3) form a ceramic matrix in the large inner surfaces of which catalytically active components are present in finely dispersed form.

Stabilizing components
Stabilizing components protect the carrier and storage components as well as the catalytically active components against excessive thermal and chemical impairments.

Honeycombs
The honeycombs to be coated in the main exhibit 10 to 250 parallel channels per square centimeter. They are usually extruded from a ceramic, e.g. cordierite, or manufactured as a special form from metal foil.

Arrangement in the exhaust-system branch
The catalysts are preferably integrated into the exhaust-gas system in such a way that the gases flow through them as uniformly as possible. Each deviation from the ideal distribution of the gases reduces the theoretically possible conversion of a catalyst such that the nonuniform flow necessitates an accordingly more powerful, i.e. larger catalyst for example.

The volume of the honeycombs, which in the technical sense is termed the catalyst volume, depends in particular on the required quantity of catalytic components which is required for the desired reduction of pollutants. A limiting factor in this respect is often the quantity which can be applied per unit of volume to the honeycomb. On the engine side the exhaust-gas back pressure of the coated honeycomb constitutes a limitation. On the vehicle side the available installation space frequently limits the catalyst volume. In Euro 6 exhaust-emissions standard diesel vehicles the total catalyst volume is often three to six times the engine displacement.

Catalytic processes for exhaust-gas treatment

Boundary conditions

Catalysts (catalytic converters) are the basis of modern processes for treating the exhaust gases of motor vehicles. Boundary conditions must be satisfied to guarantee their effectiveness in the long term. For example, the quality of the operating fluids must be matched to the respective catalysts. Leaded fuel can cause a rapid deactivation of the catalysts. Sour fuel (i.e. containing sulfur), heavy metals from the engine oil, inorganic additives, and other not completely combusted substances in the engine can cause progressive contamination of the catalysts quickly or slowly over the vehicle's service life. In this way, the efficiency of the exhaust-gas treatment system is drastically reduced in part.

In the case of the catalyst requirements imposed on the engine and vehicle applications it is also essential to adhere to certain exhaust-gas composition limits and the exhaust-gas temperature, even if critical conditions in some cases only occur on a limited basis in terms of time and space on the catalyst surface. For example, temperatures which lead to local fusing and thus destruction of a particulate filter are to be avoided during soot burn-off. This necessitates, for example, a limitation of the amount of soot to be stored.

Generally speaking, a faster progression of the catalyzed reaction can be achieved in a catalytic process with a larger quantity of active components – if the activation barrier is exceeded and if substance-transport processes do not limit the reaction speed. But even when large quantities are used the exhaust-gas temperature is always too low directly after engine starting to achieve significant pollutant conversion. The catalyst heats up with a constant exhaust-gas flow. The temperature at which 50 % of the pollutant emitted by the engine is converted is called the light-off temperature. During cold starting of a diesel engine several minutes regularly elapse before the catalysts achieve optimum operating conditions. This is counteracted by applications with engine heating measures. Electrical heating measures have also been implemented in special applications. Storage components of the catalysts converter moreover store pollutants in the cold-start phase and desorb them once the light-off temperature is exceeded. An example: Special zeolites are suitable for storing some hydrocarbons in such a way.

References

[1] https://www.arb.ca.gov/html/brochure/history.htm

Belt drives

Friction belt drives

Application

Friction belt drives are used in automobiles predominantly to drive accessories (Figure 1). These drives had previously been driven primarily by V-belts. However, the significantly higher power density required in today's applications, e.g. due to space restrictions and higher power consumption of the accessories, have resulted in these having been replaced in the meantime almost exclusively by serpentine drives with ribbed V-belts (Micro-V® belts). They are typically used to drive the alternator, the A/C compressor, the power-steering pump, but also fans, mechanical chargers, or pumps for secondary air injection.

Ribbed V-belts are also used to start engines in hybrid or in start/stop systems. Here, the conventional starter is replaced by a starter alternator, and the belt transmits the starter torque to the crankshaft to start the engine. As well as starting, supplementary functions such as additional engine acceleration (boost mode) and energy recovery by engine braking (recuperation) can be realized.

Forces and loads in the belt drive

The transmittable power is determined from (quantities, see Table 1)

$$P = (F_1 - F_2) \cdot \frac{v}{1000}\ ,\ P \text{ in kW.}$$

Table 1: Symbols and units

Quantity		Unit
F_1	Belt-run force on load side	N
F_2	Belt-run force on slack side	N
F_U	Peripheral force	N
F_{HL}	Pretensioning force or "hubload"	N
F_B	Loading of belt by wrap	N
F_C	Loading of belt by centrifugal forces in wrap angle	N
M	Transmitted torque	Nm
v	Belt speed	m/s
P	Required power transmission	kW
α_R	Groove angle, belt	°
α_S	Groove angle, pulley	°
β	Wrap angle	rad
μ	Coefficient of friction between belt rib and pulley groove	–
μ'	Coefficient of friction referred to smooth pulley	–
ω	Angular frequency	s^{-1}
L_B	Reference belt length	mm
U_B	Reference circumference	mm

Figure 1: Application example, accessory drive

1 Alternator, 2 Tension pulley,
3 Idler pulley, 4 Crankshaft,
5 Water pump,
6 Power-steering pump,
7 A/C compressor.

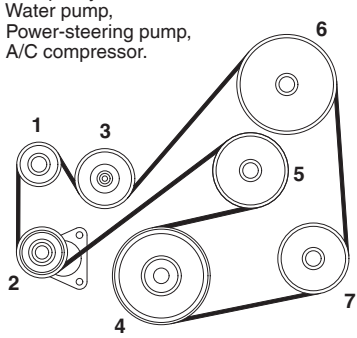
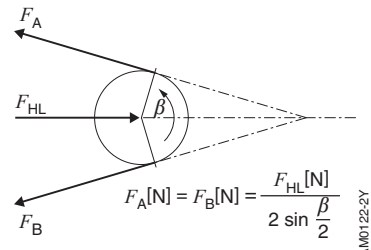


Figure 2: Forces on an unloaded belt pulley

β Wrap angle,
F_A Belt-run force, F_{HL} Hubload, F_B Belt-run force.



The transition from static friction to sliding friction is obtained from the Eytelwein equation:

$$R = \frac{F_1}{F_2} = e^{\mu' \beta},$$

where $\mu' = \frac{\mu}{\sin\frac{\alpha_s}{2}}$.

As long as the belt-run forces are in this ratio, there will be no slip when power is transmitted. Belt run refers to that section of the belt between the belt inlet and outlet areas of two adjoining belt pulleys. A typical ratio for ribbed V-belts is $R = 4$ with a wrap angle of $\beta = 180°$.

To transmit the peripheral force

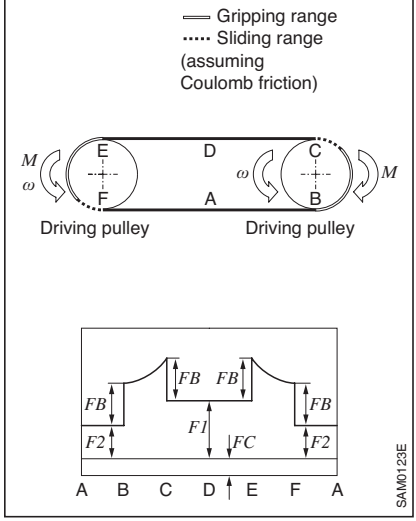
$$F_U = F_1 - F_2$$

the pretensioning force F_{HL} ("hubload") is required (Figure 2). At high rotational speeds, the centrifugal force component F_c (centrifugal force) of the belt must also be taken into consideration.

Slip occurs on account of the change in tension within the belt during alternation between load and slack runs.

In spite of a constant transmitted torque, the belt tension decreases at the driving pulley and increases at the driven pulley (Figure 3). This variation in tension causes the belt to elongate over the respective pulley. In this elongation range, the belt loses its grip on the pulley and passes into a sliding state. The belt will slip if this sliding range stretches over the entire pulley wrap. In correctly configured drives the slip is under 2 % and the ribbed V-belts operate with an efficiency of more than 96 %. This configuration must be taken into account particularly in the case of belt-driven functions, such as engine starting and engine acceleration, since driving-half and undriven-half forces are reversed for this situation.

Figure 3: Strain distribution in the belt
F_1 Force on load side,
F_2 Force on slack side,
F_B Loading of belt by wrap,
F_C Loading of belt by centrifugal forces,
M Transmitted torque,
ω Angular frequency.



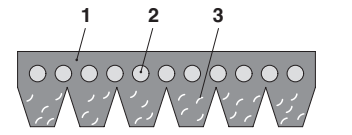
Design of ribbed V-belt

The ribbed V-belt is a composite of three components (Figure 4):
– fiber-reinforced rubber compound,
– tension members,
– backing or rubber coating.

The rubber compound forms the ribs and transmits the drive forces from the pulleys to the tension members. The main material used here is Ethylene Propylene Diene Monomer rubber (EPDM), and the rubber compound is fiber-filled for reinforcement.

The tension members accommodate the dynamic forces and transmit the drive power from the driving shaft (usually the crankshaft) to the accessories. They are usually made of nylon, polyester or aramide. These tension member materials

Figure 4: Structure of ribbed V-belt (cross-section)
1 Backing or rubber coating,
2 Tension members,
3 Rubber compound, fiber-filled.



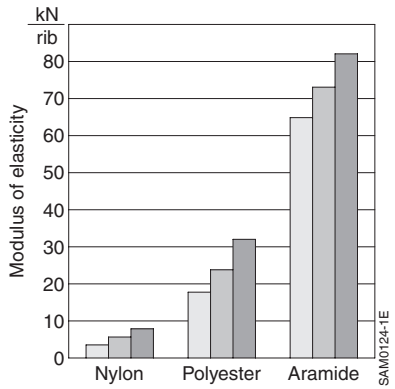
differ primarily in their very different moduli of elasticity (Figure 5).

The dynamic matching of the system can be optimized by selecting the appropriate tension member material. Tension-member materials with a high modulus of elasticity (e.g. aramide) are used in highly dynamic applications to influence the resonant frequencies of the belt drive. Depending on the choice of tension-member material, maximum dynamic belt forces of 390 N (for polyester) or up to 600 N (for aramide) per rib can be absorbed.

The back of the belt can be designed as a backing or as a rubber coating. The back of the belt forms a protective layer for the tension members. In most applications, it serves only to guide the belt quietly via the idler and tension pulleys. In some applications, however, it also serves to drive accessories which are subject to low load (e.g. water pump).

The rib profile can be manufactured by grinding, cutting or molding in a tool. The latter delivers the greatest precision in the rib geometry, helping to improve running time and noise generation. Fabric layers on the rib side additionally help to increase robustness.

Figure 5: Moduli of elasticity of different tension-member materials
The three columns for each material represent the spread.



Ribbed V-belt profile and belt pulleys
Usually the PK profile in accordance with ISO 9981 [1] is used in automobile applications (Table 2 and Figure 6).

The reference belt length L_B is determined on a two-pulley test rig under a defined pre-tension (ISO 9981). Here, the reference circumference U_B of the measurement belt pulleys is 300 mm (Figure 7). The reference belt length is then calculated according to

$$L_B = U_B + 2E .$$

Because the belt and pulley geometries in ISO 9981 are encumbered with a tol-

Table 2: PK belt profile and belt-pulley profile acc. to ISO 9981
Designation of a ribbed V-belt with six ribs of profile code PK and of reference length 800 mm: Ribbed V-belts 6PK 800.
Designation of corresponding ribbed V-belt pulley with reference diameter 90 mm: Ribbed V-belt pulley P6PK 90.

Dimensions in mm (Figure 6)	Belt	Profile grooves
Profile code	PK	K
Rib spacing s and groove spacing e	3.56	3.56
Perm. deviation tolerance for e		±0.05
Sum of perm. deviation tolerance for e		±0.30
Groove angle α_S		40° ±0.5°
Radius r_k at rib head Radius r_a at groove head	0.50	min. 0.25
Radius r_g at rib base Radius r_i at groove base	0.30	max. 0.50
Belt height h	4 to 6	
Test-pin nominal diameter d_s		2.50
2 h_s nominal dimension		0.99
2 δ		max. 1.68
Spacing f groove to thrust collar or ring		min. 2.5

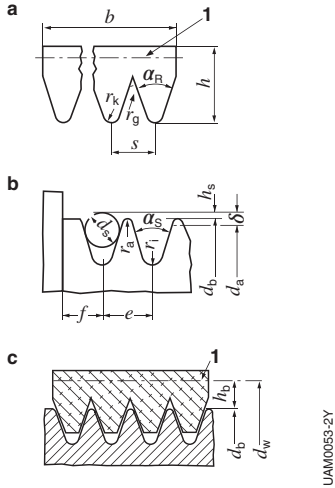
Belt width $b = n\,s$ (where n = number of ribs).
Effective diameter for profile K:
$d_w = d_b + 2\,h_b$, $h_b = 1.6$ mm.

erance band, it is absolutely essential for a detailed configuration to fall back on the characteristic values of the respective belt or pulley manufacturers.

The pulleys are made from steel, aluminum or plastic.

Figure 6: Belt profile

a) Belt cross-section,
b) Groove cross-section,
c) Determining effective diameter.
1 Position of tension member.
b Belt width,
h Belt height,
r_k Radius at rib head,
r_a Radius at groove head,
r_g Radius at rib base,
r_i Radius at groove base,
s Rib spacing,
e Groove spacing,
f Spacing groove to thrust collar or ring,
α_R Groove angle, belt,
α_S Groove angle, pulley,
h_s Allowance, test-pin nominal diameter to reference diameter,
d_s Test-pin nominal diameter,
d_a Outside diameter (diameter, groove-head profile),
d_b Reference diameter,
δ Allowance, test-pin nominal diameter to outside diameter,
d_W Effective diameter.



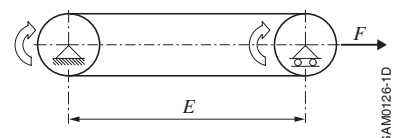
Drive system: accessory drive

The most important requirement of the accessory drive system is that it should drive all the accessories without slip. This must be guaranteed for all load states and under all ambient conditions over the entire service life of the engine. In modern engines with full drives – i.e. all the accessories are driven via an individual belt drive – maximum torques of up to 30 Nm (70 Nm for start/stop systems) and maximum outputs of 15to20 kW with all the accessories at full load are consequently transmitted via a five- or six-ribbed V-belt. The temperatures in the engine compartment are around –40 to +140 °C.

It is essential to utilize optimum system design to eliminate in particular slip noises, such as, for instance, the familiar V-belt squeaking in cold and damp weather conditions. It is also necessary to utilize design already to avoid belt noises caused by pulley misalignment. For belt-driven starting applications the shear loads to be transmitted by the ribs must most also be taken into account. Specially developed belt designs fulfill these requirements without damage caused by ribs shear off. In order to assess the service life, appropriate fatigue tests according to Wöhler must be carried out (see Wöhler fatigue tests).

Figure 7: Belt-length measurement acc. to ISO 9981

F Measurement force (pre-tension): 100 N per rib.
E Spacing of measurement belt pulleys, Reference circumference U_B of measurement belt pulleys: 300 mm.



Design criteria
Both manufacturer's own and cross-manufacturer computing programs are used to design accessory drives. Important input parameters are the arrangement of components, i.e. the drive configuration, the torque curves, the moments of inertia of the components, the torsional vibration of the crankshaft, and the belt data. With these data, it is possible to calculate and optimize the system geometry, such as, for example, belt-run lengths and wrap angle, the system natural frequencies, the slip limit angles, the belt-run forces, the belt-run vibrations, and also the service life of the belt.

Recommended minimum wrap angles

Crankshaft	150°
Alternator (I < 120 A)	120°
Alternator (I > 120 A)	160°
Power-steering pump, A/C compressor	90°
Tension pulley	60°

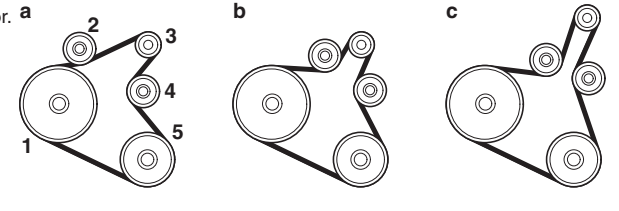
Misalignment and entry angle
In order to avoid unacceptable belt wear and noises, it is important that the entry angle of the belt into the grooved pulleys, depending on the belt design and the required service life, does not exceed 1.2 to 1.5°.

System natural frequency
The system natural frequency of the belt drive should not coincide in the idle range with the engine's ignition frequency as otherwise strong belt vibrations could occur on account of natural resonance.

Minimum diameters of pulleys and idler pulleys
In practice the smallest belt pulley is often located on the alternator in order to facilitate the high rotational speeds that are required there. Typical alternator pulleys have a diameter of 49 to 56 mm. Belt fatigue increases exponentially when very small pulleys are used; this must be taken into account when the belt is designed. Diameters not less than 60 mm are recommended for idler pulleys which generate strong belt-back deflection.

Belt-tensioning systems
Belt tension in accessory drives is today usually provided by way of automatic tension pulleys. These tension pulleys ensure that the belt is almost constantly pretensioned over the belt's service life by compensating belt elongation and belt wear. The design of the tension pulleys is significantly dictated by the space available. The system tensions are usually 40 to 70 N per rib.

Figure 8: Application examples for belt tensioner in start/stop systems
a) Two-belt-tensioner concept,
b) Pendulum belt tensioner variant 1,
c) Pendulum belt tensioner variant 2,
1 Crankshaft,
2 Tension pulley,
3 Starter-alternator unit,
4 Tension pulley,
5 Refrigerant compressor.



To monitor the dynamic response of the belt drive, common tensioning systems have a mechanical spring which works in combination with a damping unit.

Two tension pulleys are used in belt-driven start/stop systems. The tension pulleys either each have an independent spring element (two-belt-tensioner concept) or are connected to each other only by a torsion spring (pendulum belt tensioner). Start/stop belt drives with pendulum belt tensioners can be implemented with low belt pre-tension (Figure 8).

Stretch Fit drives

Elastic ribbed V-belts are also used occasionally in less complex belt drives. These use nylon as the tension-member material and do not require belt-tensioning systems. This belt type is pulled during fitting onto the pulleys by overstretching. The belt length must be designed in such a way that the belt pretensioning force is so high after fitting that, while wear and elongation over the required service life as well as the relevant ambient conditions are taken into account, sufficient pre-tension is still maintained without retensioning. Today, belt drives with up to four pulleys, i.e. up to three driven components, are possible.

Positive belt drives

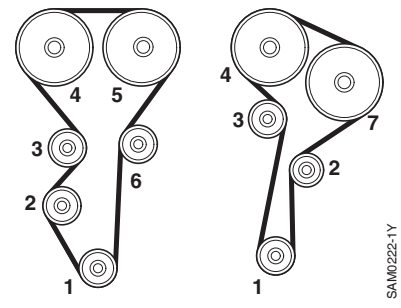
Application

Toothed belts in accordance with ISO 9010 [2] are used in timing drives to drive camshafts or the high-pressure fuel-injection pump synchronously with the crankshaft. The main advantages over competing means of drive – such as the gear drive or the chain – lie in the simplicity of the drive, the flexibility of the belt guidance, the low friction, the low noise emissions, and the capacity to compensate for high dynamic peak loads. Accessories such as oil or water pumps can be integrated into the drive or driven individually (Figure 9).

In modern applications, in contrast to earlier applications, it is possible in many cases to dispense with a replacement interval because of the use of innovative belt technologies and optimized design of the overall system. The low weight and the low costs also militate for the use of a toothed belt.

Figure 9: Application examples for toothed-belt drives

a) In-line gasoline engine,
b) In-line diesel engine.
1 Crankshaft,
2 Water pump,
3 Tension pulley,
4 Camshaft 1,
5 Camshaft 2,
6 Idler pulley,
7 High-pressure fuel-injection pump.

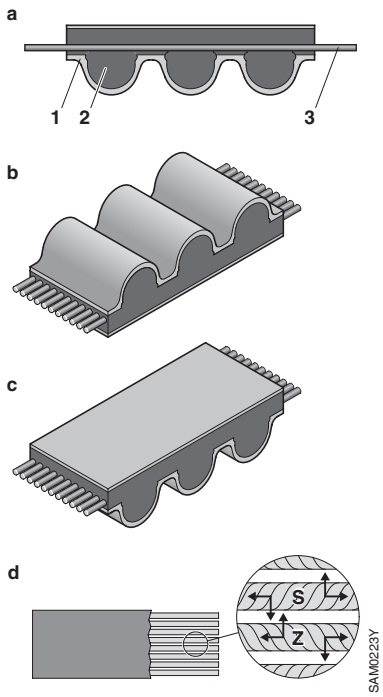


Forces and loads in the belt drive

In positive belt drives the torque is transmitted by the meshing of the toothed-belt teeth and a special toothed-belt pulley from the crankshaft to the toothed belt and from the toothed belt to the output components (e.g. camshaft, high-pressure fuel-injection pump or water pump). The actual power transmission is performed by the tension members embedded in the toothed belt. The toothing renders slip impossible, which means that toothed-belt drives can be used for synchronization.

Figure 10: Structure and characteristics of toothed belt

a) Longitudinal section,
b) Overhead view, tooth side,
c) Overhead view, back side,
d) Overhead view with depiction of tension members twisted in S- and Z-shapes.
1 Nylon fabric,
2 Rubber compound,
3 Tension member.



Structure of toothed belt

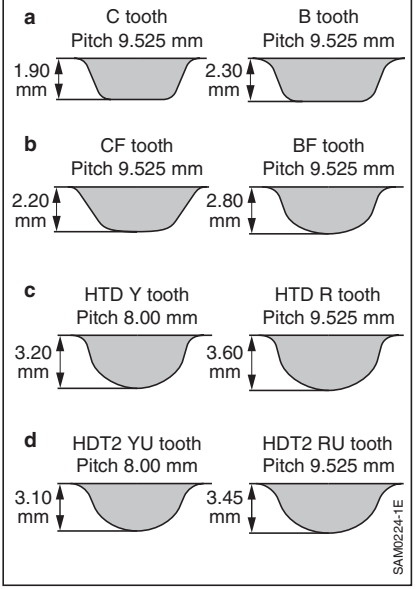
The toothed belt is a composite of three components (Figure 10):
– nylon fabric,
– rubber compound, and
– tension member.

The fabric is made of high-strength nylon and has an abrasion-proof and wear-resistant coating. It protects the rubber teeth against wear and from being shorn off, and improves the running performance.

The rubber compound consists of a high-strength polymer, and encloses the tension member on both sides. Polychloroprene (CR) was used in the original applications. The high demands placed on resistance to temperature and aging and dynamic strength means that today only Hydrogenated Nitrile Butadiene Rubber

Figure 11: Development of toothed-belt profiles

a) Power-Grip trapezoidal profile,
b) Power-Grip Power Function profile,
c) Power-Grip HTD profile,
d) Power-Grip HTD2 profile.



materials (HNBR) are still used in today's motor vehicles.

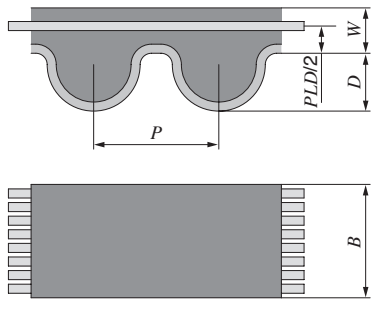
In some high-load applications belts with backing can also be used for increased reinforcement. This fabric reduces the risk of the belt running in the axial direction. It also improves belt service life in very cold ambient conditions (down to –40 °C), for example when the engine is started.

The tension member consists of twisted glass fibers – a material which is characterized by high tensile strength together with a high willingness to bend. As a result of the production process, the tension members are arranged in a spiral shape in the belt in the effective line, and twisted in pairs in S- and Z-shapes (Figure 10d). This ensures that the belt is subject to predominantly neutral running.

Toothed-belt profiles

The first camshaft belts were based on the classic Power Grip trapezoidal-tooth shape (Figure 11a), as was already known in industrial applications. On account of the increasing demands with regard to load transmission, jump protection and noise, circular-arc-like profiles (e.g. Power Grip HTD 2, High Torque Drive, Figure 11c and 11d) are almost exclusively used today. Compared with trapezoidal teeth, the forces with circular profiles are introduced more uniformly into the tooth, and stress concentrations are thereby avoided.

Figure 12: Characteristics, toothed belt
a) Longitudinal section, b) Overhead view.
P Pitch, *D* Addendum, *W* Rib thickness, *B* Width, *PLD*/2 Effective line distance (half the distance from outside diameter to effective line)



The pitch (Figure 12) is, in most cases, 9.525 mm for diesel engines, and 8.00 mm for gasoline engines. Higher forces can be transmitted with a larger pitch, while a smaller pitch has advantages with regard to noise and space.

Double-sided toothed belts can be used for applications where the direction of rotation is reversed (e.g. balancer-shaft drives).

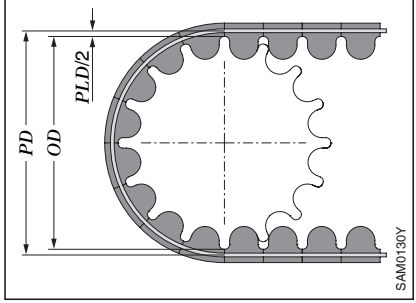
Toothed wheels

The associated toothed wheels are specified in ISO 9011 [3]. For the toothed wheel, the profile must be determined as a function of the diameter. The effective diameter *PD* is obtained from the number of teeth and the pitch, the outside diameter of the toothed wheel is accordingly reduced by the *PLD* (see characteristics in Figure 12 and Figure 13).

Drive system: toothed belt

The most important requirement of the toothed-belt drive system is that it should synchronize the camshaft in relation to the crankshaft over the service life of the engine. This is an important criterion for adhering to the consumption and emission values. It is possible to keep the elongation of the toothed belt under 0.1 % of the belt length by selecting the appropriate materials for the toothed belt, and by using an automatic tensioning system and optimized system dynamics. This results in 4-cylinder engines in a timing deviation of 1 to 1.5° referred to the crankshaft.

Figure 13: Characteristics, toothed wheel
PD Effective diameter, *OD* Outside diameter, *PLD*/2 Effective line distance.



The service-life requirements are currently in the range of 240,000 to 300,000 km, while the temperature requirements range from –40 °C in the cold state up to 150 °C engine-compartment temperature. Conspicuous noises caused for example by meshing are perceived to be offending and are therefore not acceptable. The efficiency of a toothed-belt drive is approx. 99 %. New materials and manufacturing processes today permit the use of toothed belts in an oil environment to drive oil pumps or to control the timing of camshafts. For the latter, it is possible to dispense with leading the camshaft out of the engine block and hence with the customary shaft seal on the main bearing. The advantages of this technology are the potential friction and noise reductions. The oil environment has not only a friction-reducing, but also a damping effect. This in turn has a positive impact, particularly for highly dynamic downsizing engines.

Design criteria

Both manufacturer's own and cross-manufacturer computing programs are used to design timing drives.

Important input parameters are the arrangement of components, i.e. the drive configuration, the torque curves of the components, and calculated from these the dynamic peripheral forces, as well as the belt data. With these data, it is possible to calculate and optimize the system geometry, such as, for example, belt-run lengths and wrap angles, and the system dynamics.

Recommended minimum wrap angles

Crankshaft	150°
Camshaft, high-pressure pump	100°
Accessory wheel	90°
Automatic tension pulley	50°
Idler pulley	30°
Damping pulley	10°

Periodic meshing

To avoid uneven belt wear, it is important to prevent the same belt teeth from engaging in the same wheel gaps. To assess this periodic meshing the tooth ratio of the number of teeth of the belt and the belt pulleys and the number of engine cylinders or the number of pistons in the high-pressure pump are considered.

Z	Tooth ratio
AM	Number of engine cylinders
AHD	Number of pistons in the high-pressure pump
ZR	Number of teeth on the belt
ZSNW	Number of teeth on the camshaft disk
ZSHD	Number of teeth on the high-pressure pump disk

$Z = ZR \cdot AM/ZSNW$
$Z = ZR \cdot AHD/ZSHD$

Periodic meshing does not occur when the following decimal values are avoided for the tooth ratio:

Z.0; Z.25; Z.33; Z.5; Z.66; Z.75.

Belt-run lengths
In order to avoid resonance noises at idle speed, it is important to ensure that free belt-run lengths are not in the area of 75 mm and 130 mm.

Minimum number of teeth and minimum diameters of toothed wheels and idler pulleys

Pitch 9.525 mm	17 teeth
Pitch 8.00 mm	17 teeth
Non-toothed idler pulleys	dia. 50 mm
Damping pulleys	dia. 28.5 mm

Tolerances of toothed wheels and idler pulleys
Concentricity, lateral running:
dia. ≤ 100 mm 0.1 mm
dia. > 100 mm 0.001 mm per mm dia.

Conicity of outside diameter
≤ 0.001 mm per mm wheel thickness.

Parallelism of bore to toothing
≤ 0.001 mm per mm wheel thickness.

Surface roughness
$R_a \leq 1.6$ µm.

Pitch error
– dia. ≤100 mm:
  ±0,03 mm gap/gap,
  max. 0,10 mm over 90°.
– dia. > 100 mm:
  ± 0.03 mm gap/gap,
  max. 0.13 mm over 90°.

Axial guidance
A toothed belt must be guided at least at one wheel on both sides with flanged wheels to prevent it running off the belt run. Essentially, it is important to ensure that toothed wheels with flanged wheels are exactly flush with the other wheels so that the belt is not deflected from its course. Limit values for belt-edge wear caused by excessively high run-off forces must be taken into consideration.

Belt-tensioning systems
The required constantly high belt tension, and compensation of the tension increase over temperature and belt elongation are usually provided in today's timing drives by tension pulleys. The design of the tension pulleys is significantly dictated by the space available. The most widely used system is the mechanical, friction-damped compact tensioner.

Essential design parameters for monitoring the system dynamics are the spring moment and the damping characteristic. Hydraulic tension pulleys are also used in some applications in the event of very high dynamic forces in the toothed-belt drive. These are characterized by their asymmetrical damping performance.

Tensioning systems can also be used in an oil environment if an appropriate adaptation of the damping level is taken into consideration.

References
[1] ISO 9981: Belt drives – Pulleys and V-ribbed belts for the automotive industry – PK profile: Dimensions.
[2] ISO 9010: Synchronous belt drives – Automotive belts.
[3] ISO 9011: Synchronous belt drives – Automotive pulleys.

Chain drives

Overview

Camshafts are used to control the gas exchange valves; these camshafts open and close the valves by means of direct actuation via bucket tappets or lever assemblies. Camshafts are driven on modern overhead valve engines, i.e. overhead camshafts, with the aid of a belt drive, while side-mounted camshafts are connected via a geartrain to the crankshaft pinion. Toothed belts, roller- or sleeve-type chains, and tooth-type chains are used to drive overhead camshafts. For fast-running assemblies – such as motor-racing engines for example – the drive is also effected with spur gears.

The most important criteria in deciding on the type of drive are costs, space, ease of maintenance, service life, and noise buildup. The great advantage of chain drives over toothed-belt drives is that they are completely maintenance-free over the entire service life of the engine, while toothed belts, depending on the application, have to be retensioned or changed according to the maintenance interval.

In addition to the camshaft, timing drives in modern engines frequently drive other assemblies, such as the oil pump and the high-pressure pump for gasoline direct injection or for common rail (Figure 1).

As both camshaft and crankshaft rotate nonuniformly and are exposed to torsional vibrations, the drive is subject to highly complex dynamic stresses. Furthermore, the torque demand of the high-pressure pump is subject to very strong periodic fluctuations, and thus generates further excitation of the timing-chain drive.

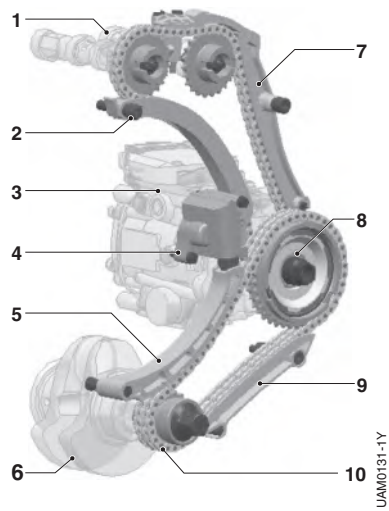
Timing chains

Designs of steel-link chains
In the case of standard chains, a distinction is made between roller- and sleeve-type chains. There are also simplex and duplex chains. The tooth-type chain is a special type of chain.

Roller-type chain
A steel-link chain is composed of inner and outer links. The inner link of a roller-type chain consists for example or two inner plates and two sleeves pressed into the plate eyes (Figure 2a). The outer link consists of two outer plates and two link pins which connect the inner and outer links. In a roller-type chain a roller is fitted over the sleeve.

The rotating rollers of a roller-type chain located over the sleeves roll with little friction on the tooth flanks of the sprocket

Figure 1: Timing-chain drive
1 Camshafts, 2 Tensioning rail,
3 Fuel-injection pump,
4 Chain-tensioner system, 5 Tensioning rail,
6 Crankshaft, 7 Guide rail,
8 Intermediate-shaft sprocket group,
9 Guide rail,
10 Crankshaft sprocket.



such that a different point on the circumference is always contacted. The lubricant between the rollers and sleeves helps to muffle noise and dampen shocks.

Sleeve-type chain
In the case of a sleeve-type chain (Figure 2b), the tooth flanks of the sprocket constantly touch the fixed sleeves at the same point such that the sleeves are subjected to an additional load. Perfect lubrication is therefore particularly important for these drives.

While having identical pitch – the chain pitch defines the distance from pin center to pin center – and force at break, sleeve-type chains have a larger joint surface than the corresponding roller chains, since the omission of the roller means that the pin diameter can be enlarged. A larger joint surface results in a lower joint-surface pressure and thus lower wear in the joints and a longer service life for the chain timing drive.

Sleeve-type chains have proven useful in camshaft drives subject to high loads in diesel engines, because increased wear resistance is called for here on account of the increased entry of soot into the engine oil.

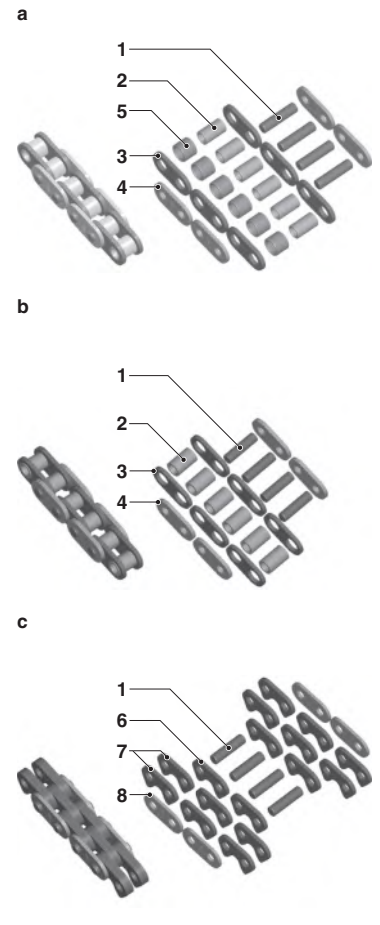
Tooth-type chain
A special design of steel-link chain is the tooth-type chain. On a tooth-type chain (Figure 2c), the plates are set out in such a way that they can transmit the force between the chain and the sprocket, whereas, in the case of roller- or sleeve-type chains, the connection with the sprocket is established via pin, sleeve, or roller.

Tooth-type chains can be built to virtually any desired width without the need for fundamental structural change. Guide plates, which are located either in the middle or on the outside (on both sides), are fitted to prevent the sprocket from running off.

A further variant of the tooth-type chain is the design that is toothed on both sides, which can then be used like a roller- and sleeve-type chain.

Sleeve-/tooth-type chain
Combining the advantages of a sleeve-type chain and a tooth-type chain creates a new type of chain for timing drives, a sleeve-/tooth-type chain (Figure 3). On the one hand, its special arrangement of tooth plates permits a smooth run into the sprocket; on the other hand, the use of sleeves reduces the wear on pins and

Figure 2: Chain designs
a) Roller-type chain,
b) Sleeve-type chain,
c) Tooth-type chain.
1 Link pin, 2 Sleeve, 3 Inner plate,
4 Outer plate, 5 Roller, 6 Tooth center plate,
7 Tooth inner plate, 8 Guide plate.



plates. This chain variant is recommended in all applications where excellent wear resistance and good acoustic and dynamic performance are important.

Selecting chain types

When choosing the chain types (construction and pitch), it is important, while adhering to the maximum sprocket diameter, to aim for numbers of teeth of more than 18 teeth in order to reduce dynamic influences by the polygon effect. The polygon effect occurs when a traction mechanism (chain) is positively driven by a drive wheel. With the sprocket at constant angular velocity the chain is accelerated and decelerated and consequently dynamically loaded.

Decades of experience have shown that some dimensions for roller-, sleeve- and tooth-type chains are particularly suitable for timing drives. These are 3/8" sleeve-type chains for diesel engines and 8 mm sleeve-type chains and 8 mm roller-type chains for gasoline engines. If in the development of a gasoline engine particular demands are made with regard to engine acoustics, 8 mm or 6.35 mm tooth-type chains should be used.

Requirements of chains

Four important factors determine the usage characteristics of timing chains: ultimate tensile strength, fatigue strength, resistance to wear, and acoustic behavior. A break may be caused by the static or dynamic load at break being exceeded. Especially in the case of timing drives, no uniform load will be encountered such that the fatigue strength of the chain is the strength-limiting quantity. The chain is subjected to a dynamic load as a result of the pulsating torques of the camshafts, the high-pressure pump, the rotational non-uniformity of the crankshaft, and the pulsating axial chain force caused by the polygon effect. Here, the safety limit to the fatigue strength of the chain must not be exceeded, since the number of such load alternations during an engine's service life is in each case greater than 10^8 load alternations (Figure 4).

In today's engines with precise timing and minimal clearance between piston and valve, minimal elongations due to chain wear can be achieved. Length increases caused by wear of just 0.2 to 0.5 % of the chain length for an engine mileage of up to 350,000 km are ensured today by the optimization of the pin/sleeve joint.

In particular direct-injection turbodiesel engines but also the latest models of direct-injection gasoline engines require timing chains that exhibit the highest levels of wear resistance.

To improve the wear resistance, the link pins are coated in a final operation with a layer of hard material, either chromium carbide or chromium nitride.

Figure 3: Sleeve-/tooth-type chain

1 Sleeve,
2 Tooth plate,
3 Guide plate,
4 Pin.

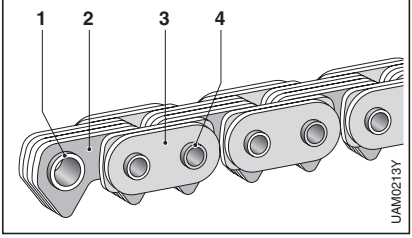
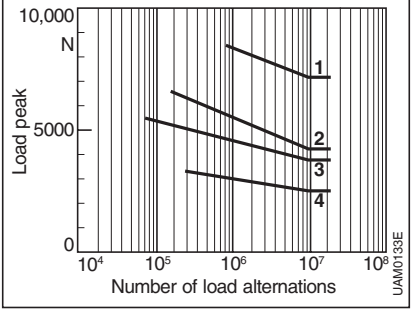


Figure 4: Fatigue-strength results for sleeve- and roller-type chains

1 Sleeve-type chain for DI applications,
2 Sleeve-type chain for prechamber diesel engines,
3 Roller-type chain for high-performance timing drives in gasoline engines,
4 Roller-type chain for oil-pump drives.



Design criteria for chain timing drives

A chain timing drive is with mass, rigidity, and damping an oscillatory system with several degrees of freedom. If subjected to corresponding excitation for example by the camshaft, crankshaft or high-pressure pump, this can, on account of interactions, cause resonance effects which lead to extreme timing-drive load.

A range of design criteria must be taken into account in the development of chain timing drives. These are:

– Chain length with an even number of links.
– Free run length (unguided chain section) if possible three to five chain links.
– Large radii at tensioning and guide rails for reducing friction.
– Inward tensioning direction of the chain tensioner.
– Floating mounting of the tensioning and guide rails.

Computational check of chain timing drives

Layout drawings are taken as the basis to create a chain-drive model which is used for the dynamic computation of timing drives. The timing drive is described by mass parameters, chain stiffness, and damping values.

Tensioning and guide rails are modeled as elastic bodies, the tensioning element is depicted with all the function elements. The simulation calculation delivers the necessary torsional-vibration angle deviations and all the force and moment characteristics.

Friction reduction of chain timing drives

To reduce the CO_2 emissions from an internal-combustion engine, every savings potential must be utilized to reduce fuel consumption. In this way, a friction-optimized timing-drive design can also make a contribution to reducing the CO_2 emissions.

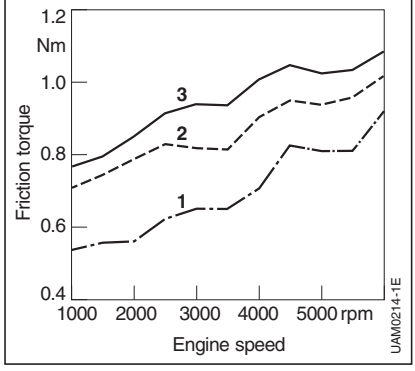
The decisive factor is the optimal layout of the chain line. Avoiding pronounced curvatures of the tensioning and guide rails reduces the normal forces and hence reduces the friction.

Chain-plate quality is of crucial importance to the frictional behavior of timing chains. The manufacture of chain plates involves the use of different blanking processes which create textures on the surface based on the manufacturing process used. Chains with precision-blanked plates deliver the best values.

Due to their internal friction and the contact of pins and plates, even timing chains of the most important types exhibit different frictional behavior. Figure 5 clearly shows the disadvantages of the tooth-type chain with identical overall width.

Figure 5: Friction torque of different chain designs

Chain pitch is the same for all designs at 8 mm.
1 Sleeve-type chain,
2 Roller-type chain,
3 Tooth-type chain.



Sprockets

The tooth shape of the sprockets is standardized for roller-, sleeve- and tooth-type chains (DIN ISO 606, [1]). DIN permits considerable degrees of freedom in the precision design of the sprocket toothing such that a tooth-space shape can be created between a minimum space and a maximum space. As a rule, sprockets with maximum tooth-space shape are used. This design permits, as a result of the low addendum and the larger tooth-space opening, the chain to enter and exit undisturbed even at higher chain speeds.

In fashioning the tooth shape of tooth-type chains each manufacturer applies its own design philosophy (e.g. noise- or wear-optimized) such that these chains do not conform to the standard. The design of the sprockets is dependent on the space conditions, the operating conditions, and the power transmission of the chain drive (Figure 6).

Sprockets made of carbon steel (e.g. Ck45) and alloyed steels (e.g. 16MnCr5) and wheels made of sintered materials (e.g. Sint D11) are used. Precision-blanked and machine-cut wheels with the heat treatment appropriate for the material are used.

Chain-tensioning and chain-guide elements

By using permanently acting clamping and guide elements (Figure 1) which are precisely matched to the engine in question, it is possible to optimize the chain drive in such a way that its service life corresponds to that of the engine.

Chain tensioner

The chain tensioner (Figure 7), mechanically or hydraulically acting, assumes a range of tasks in the timing drive. On the one hand, the timing chain is pre-tensioned in all operating conditions on the loose side under a defined load, even in the event of wear elongation that occurs during operation. Slack run refers to the non-pulled, as it were load-free chain section. A damping element – either friction or viscous damping – reduces vibrations to an acceptable level.

Mechanical chain tensioners without additional hydraulic damping are used, as

Figure 6: Sprockets
a) Sintered crankshaft gear,
b) Precision-blanked camshaft gear,
c) Sprocket with sealing surface,
d) Two-track sprocket for duplex chain.

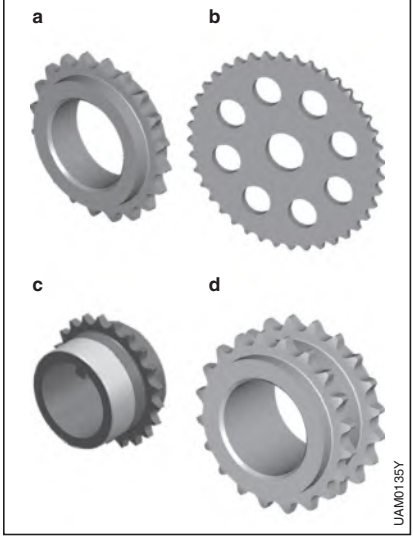


Figure 7: Types of chain tensioners
a) Mechanical chain tensioner,
b) Hydraulic chain tensioner.

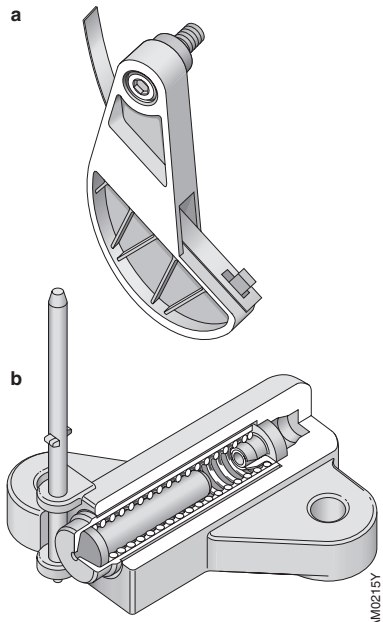
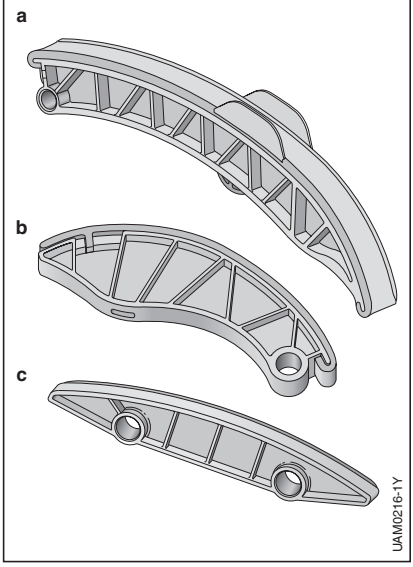


Figure 8: Types of guide and tensioning rails
a) Tensioning rail, two-component type,
b) Tensioning rail, two-component type, with lubricating duct,
c) Guide rail, one-component type.



a rule, in oil-pump drives which are subject to low load. In special cases, this mechanical chain tensioner can even be omitted completely.

Tensioning and guide rails

Simple rails made of plastic or metal (aluminum or sheet steel) with plastic covering and which are flat or bent - depending on the chain route - are sometimes used as tensioning and guide elements. In the case of newer types, the rails are for the most part inexpensively injection-molded from plastic and manufactured as a two-component design (base carrier and sliding lining).

Here, in the case of the tensioning rail, a friction lining of glass-fiber-free polyamide is sprayed or clipped onto a carrier of heat-resistant polyamide with 30 to 50 % glass-fiber content. The sliding rails are usually designed in plastic as one-component rails, and are used for chain guidance. Figure 8 shows some types of tensioning and guide rails.

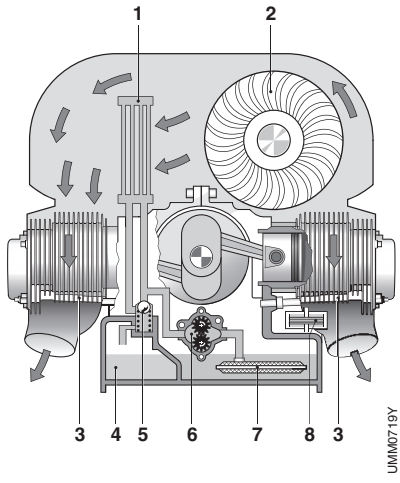
References

[1] DIN ISO 606: Short-pitch precision roller and bush chains, attachments and associated chain sprockets (ISO 606:2004).

Engine cooling

The process of converting any type of energy into mechanical energy to drive a vehicle creates heat as a waste product, whether it be by friction of different engine components against each other or by the combustion of fuel in the engine. In order to optimally utilize the conversion of the chemical energy of fuel into kinetic energy, this process must take place under controlled conditions and, to protect the engine and its components, also under controlled temperatures. The waste heat that is generated must be reliably dissipated to atmosphere under all operating and ambient conditions.

Figure 1: Principle of air cooling
1 Oil cooler,
2 Cooling-air fan,
3 Cylinder with cooling fins,
4 Oil pan,
5 Pressure-relief valve,
6 Oil pump (gear pump),
7 Lube-oil filter,
8 Thermostat.



Air cooling

Design and operating principle

Cooling air is routed by dynamic pressure or a fan around the external walls of the cylinder casing (Figure 1). These walls are finned in order to achieve better cooling results thanks to the increased surface area. The cooling-air quantity can be controlled for example by specifically restricting its flow at the vehicle's cooling-air inlets, which is achieved for example with thermostat-controlled flaps or by means of the rotational speed of a load- or temperature-dependent fan.

Heat absorbed by the engine oil is dissipated via cooling fins on the oil pan which are also located in the air stream.

Advantages and disadvantages of air cooling

The advantages of air cooling are simple and cost-effective construction, reliable operation, and lower weight when compared with water-cooling systems.

The disadvantages are higher noise generation, lower temperature consistency of all engine parts, and poorer thermal behavior at high specific power outputs.

Air-cooled engines struggle to comply with the currently required exhaust-emission limits.

Applications

Today, air cooling is mainly used for motorcycle and aircraft engines and in special applications. Extra-small engines in model-making applications also use air cooling.

The expected operating conditions for a vehicle can only dictate which cooling principle is chosen. In Arctic regions, where temperatures down to −50 °C are experienced, water-cooled engines cannot be used owing to the risk of the water freezing; for this reason air cooling must be used in these regions.

Water cooling

Design and operating principle

Cooling circuit

In contrast to air cooling, with water cooling the waste heat from the engine is not dissipated directly to the ambient air. Instead, the waste heat is absorbed by the coolant and transported via this medium to the main coolant radiator (Figure 2). Here, heat is exchanged between the coolant and the ambient air. The cooling-air flow rate is controlled by a fan, and the coolant is circulated in the circuit by a coolant pump.

Coolant cooling, also called water cooling, has become the standard in both passenger cars and commercial vehicles.

Coolant

Coolants are a mixture of water, antifreeze (mostly ethylene glycol), and corrosion inhibitors. The inhibitors prevent the components through which coolant flows from corroding (see Coolants). Adding antifreeze in concentrations of 30 to 50 % firstly stops the coolant from freezing even at temperatures down to – 25 °C and secondly raises the coolant boiling temperature to allow coolant temperatures of up to 120 °C at a pressure of 1.4 bar in passenger cars.

Radiator designs

The cores of the coolant radiators in modern passenger cars are almost exclusively made of aluminum. Aluminum radiators are also being used to an ever-increasing extent in a wide range of commercial vehicles and trucks throughout the world. There are two assembly variants: brazed radiators and mechanically joined or assembled radiators.

Flat-tube/corrugated-fin systems

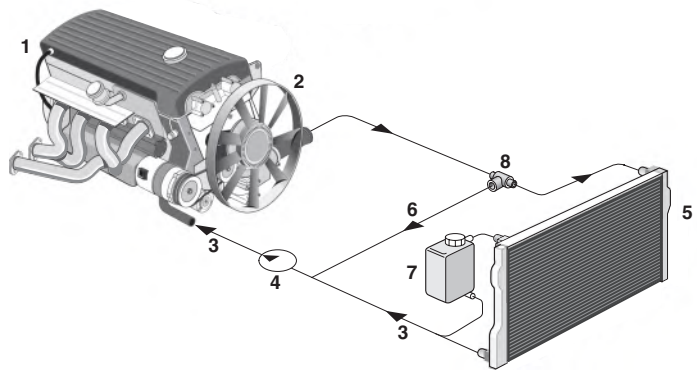
For cooling high-performance engines, or when space is limited, the best solution is a brazed flat-tube and corrugated-fin radiator layout with minimized aerodynamic resistance on the air-intake side.

Construction

The radiator core consists of series of flat tubes with corrugated fins located in between (Figure 3). The flat tubes lead into the openings in the radiator base. During the manufacturing process the tips of the corrugated fins are soldered to the

Figure 2: Passenger-car cooling system

1 Engine,
2 Fan,
3 Coolant line (arrow indicates direction of flow),
4 Coolant pump,
5 Main coolant radiator (flat-tube/corrugated-fin system),
6 Bypass line (secondary cooling circuit),
7 Expansion tank with pressure-relief valve (integrated in cap),
8 Thermostat.



flat tubes and the flat tubes to the radiator base in an oven. The solder material is applied as a layer to the materials.

The radiator tank ensures that the coolant is distributed throughout the core. These tanks are made of fiberglass-reinforced polyamides, and are injection-molded with all connections and mountings in a single unit. They are flange-mounted to the radiator core and sealed by an integrated elastomer sealing element.

Operating principle

The coolant is routed through the radiator tank to the radiator core. It flows through the flat tubes before emerging again from the opposite radiator tank. The hot coolant in the flat tubes dissipates heat to the corrugated fins, while the cooling air flowing through the radiator core cools the corrugated fins. The soldered connections guarantee a good heat transfer from the tubes to the fins. Gills in the corrugated fins, which are brought together in gill fields, bring about additional swirling of the cooling air. This results in improved heat dissipation.

Radiator tanks are designed in such a way as to distribute the coolant as uniformly as possible to all the flat tubes. The tubes have minimal depth (approx. 2 mm with wall thicknesses of 0.2 to 0.3 mm) so that the full coolant flows past the tube walls as closely as possible, thereby ensuring a good heat transfer to the fins.

Tube/fin systems

The less expensive, mechanically assembled tube/fin system is generally employed for applications with less powerful engines or when more space is available.

Construction

When the radiator is assembled mechanically, the cooling grid is formed by mounting stamped cooling fins around round, oval or flat-oval tubes (Figure 4). The distance between the cooling fins is maintained by trapezoidal cutouts. In the manufacturing process the tubes are

Figure 3: Flat-tube/corrugated-fin system
1 Radiator tank,
2 Radiator base,
3 Gasket,
4 Flat tube,
5 Corrugated fins,
6 Gill zones.

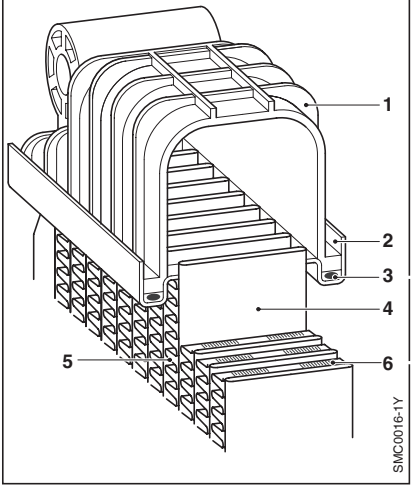
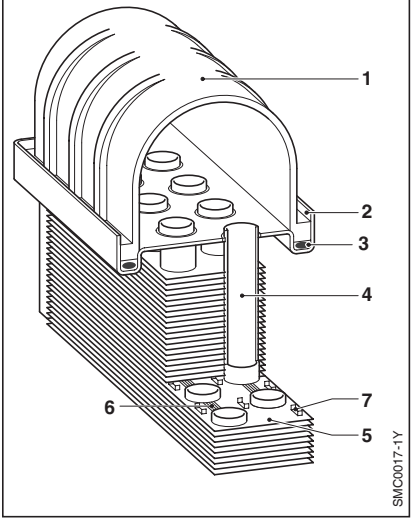


Figure 4: Tube/fin system
1 Radiator tank,
2 Radiator base,
3 Gasket,
4 Round tube,
5 Cooling fin,
6 Gill zones,
7 Turbulators.



initially inserted loosely into the cooling fins and flared with a tool. This creates an internal connection between the tubes and the fins. The connection is established exclusively by the mechanical flaring of the tube. However, the heat transfer between the tubes and the fins is much worse than in the soldered systems.

Operating principle
The cooling air flows through the radiator core between the cooling fins. These are corrugated or slotted at right angles to the direction of air flow. They are – like the corrugated fins in soldered radiators – also fitted with turbulators and gill zones, inducing air turbulence and improving the cooling effect.

Options for increasing radiator performance
Gills and a larger number of corrugations are integrated into the cooling fins on the cooling-air side to enhance thermal transfer. The gills induce turbulence in the cooling air flowing through, and the larger number of corrugations produces a larger surface area. These measures increase the quantity of heat that can be dissipated.

Further measures to increase cooling efficiency include the use of tubes with the lowest possible width and wall thickness, and turbulators on the coolant side, provided the attendant pressure losses remain within acceptable limits. The negative effect of these performance-enhancing measures is the higher pressure drop on both the cooling-air and coolant sides, which must be compensated by a higher-capacity fan and on the coolant side by a more powerful coolant pump.

Coolant expansion tank
Mounted at the highest point of the cooling circuit is the airtight expansion tank; this tank has a filler neck through which coolant is poured into the cooling system. When the engine is running the expansion tank takes up excess coolant that expands due to heating. Where necessary, it returns it to the cooling circuit. As the coolant heats up, so too the system pressure in the cooling system increases. This causes the coolant boiling point to rise.

The expansion tank's air volume must be large enough to absorb the coolant's thermal expansion during rapid pressure buildup, and prevent the coolant from boiling over under the permitted operating conditions. At excessively high operating temperatures and thus also excessively high operating pressures the expansion tank protects the cooling circuit by means of a pressure-relief valve.

The expansion tank also provides a reliable escape channel for pressurized gases, preventing cavitation of the kind that tends to occur on the intake side of the coolant pump.

Expansion tanks are injection-molded in plastic (generally polypropylene), although simple designs can also be blow-molded to shape. A system of hoses is normally used to connect the expansion tank to the cooling system. The expansion tank is mounted in the engine compartment at the location representing the highest point in the cooling system to ensure air is expelled effectively. In some cases, the expansion tank can form a single unit with the radiator tank, or the two can be joined by means of a flange or plug connection.

The position and shape of the filler neck can be used to limit capacity, thus preventing overfilling. An electronic level sensor can be fitted to monitor the level of the coolant. The coolant level can also be monitored by manufacturing the expansion tank completely or partially in natural-color, transparent plastic with molded level markers. However, colorless polypropylene is sensitive to ultraviolet rays. For this reason, the transparent part of the expansion tank should not be exposed to direct sunlight.

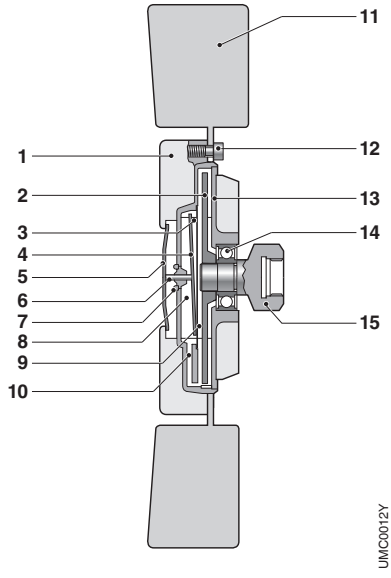
Cooling-air fan
Design
Since motor vehicles also require substantial cooling capacity at low speeds, force-air ventilation is required for the radiator. Injection-molded plastic fans with drive-power ratings extending up to 30 kW are now used as standard in commercial vehicles. A fan of this type is driven by a mechanical coupling with the internal-combustion engine, for example via a belt drive. The fan can also be mounted directly on the crankshaft.

Single-piece injection-molded plastic fans are generally used in passengers cars; these are usually driven by DC brush motors or brushless DC motors. These motors are installed in the fan hub. The electrical drive power is around 400 W in the small-car segment, and extends up to 1 kW in the luxury-class segment and in offroad vehicles. Although blade design and arrangement can be selected to provide relatively quiet running, these fans have considerable noise levels at high rotation speeds.

In some passenger-car applications, in particular offroad vehicles with very high engine power outputs in combination with equipment options for operation in hot climates and equipment variants for diesel engines and air-conditioning systems, the capacity of the electric drive is no longer sufficient to provide the volume of air necessary to ensure effective cooling. Fan-output figures in excess of 1 kW can only be achieved when the fan is driven mechanically by the engine. However, this is only possible with a longitudinally installed engine.

Figure 5: Bimetall-controlled viscous clutch

1 Cover with cooling fins,
2 Primary disk,
3 Valve bore,
4 Valve lever,
5 Thermal bimetallic strip,
6 Operating pin,
7 Seal,
8 Supply chamber,
9 Working chamber,
10 Return bore,
11 Fan,
12 Mounting bolt,
13 Basic body,
14 Ball bearing,
15 Flanged shaft.



Control of electrically driven fans
Depending on vehicle and operating conditions, the unassisted air stream can provide sufficient cooling up to 95 % of the time. It is thus possible to economize on energy that would otherwise have to provide the power to drive the fan. For this purpose, electric fans use a multistage or continuous control system that specifically adapts the operating periods and speed of the fan to the required cooling capacity. A multistage control system may consist of relays and series resistors, while continuous variable control necessitates the use of power electronics. Electric thermostatic switches or the engine control unit supply the input signals for the control system.

Drive of mechanical fans
The fluid-friction clutch (viscous clutch) is a mechanical-drive arrangement of proven effectiveness for use in commercial vehicles. It essentially consists of the following assemblies (Figure 5):
– Input or primary section (flanged shaft and primary disk)
– Output or secondary section (cover, basic body)
– Control facility for controlling and regulating the oil filling (working fluid)

The secondary section of the viscous clutch is divided into a working chamber and a supply chamber. When the viscous clutch is not cut in, only a small amount of working fluid is in the working chamber, such that only a very small amount of torque can be transmitted due to the high degree of slip between primary disk and secondary section. The viscous working fluid is a silicone oil.

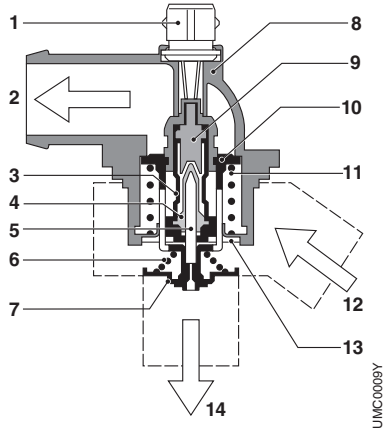
As the temperature increases, the bulging of the bimetallic strip causes the operating pin to open the valve. Silicone oil flows from the supply chamber into the working chamber. This reduces the slip between primary disk and secondary section – in turn increasing the fan speed. The fan speed and thus the cooling capacity are infinitely increased.

As the temperature decreases, the bimetallic strip cools and causes the operating pin to close the valve slowly. The silicone oil flows through the pump body back into the supply chamber. The fan speed decreases. The amount of silicone oil located in the working chamber determines the output power transmitted by the clutch and therefore the speed of the fan.

A distinction is made between types of viscous clutch, based on how the valve is actuated: Firstly, as already described, the temperature-dependent, self-governing clutch which varies its speed infinitely by means of a bimetallic element, an operating pin, and a valve lever. The controlled variable is the temperature of the air leaving the radiator, and thus indirectly the temperature of the coolant. The electrically activated clutch is now used in 95 % of heavy-duty commercial vehicles. In this clutch type the valve lever is electromagnetically actuated and thus the oil quantity in the working chamber is regulated. Instead of just one controlled variable, a wide range of input variables is used for control purposes. These are usually the temperature limits of the various cooling media.

Figure 6: Map-controlled thermostat

1 Connector, 2 Connection to radiator,
3 Housing of working element,
4 Elastomer insert,
5 Plunger,
6 Bypass spring,
7 Bypass valve,
8 Housing,
9 Heating resistor,
10 Double-acting disk valve (main valve),
11 Main disc spring,
12 Connection from engine,
13 Tie bar,
14 Connection to engine (bypass).



Regulation of coolant temperature

A motor vehicle's engine operates in a very wide range of climatic conditions and with major fluctuations in engine load. The consequences of this are significant variations in coolant and engine temperatures, which in turn result in increased engine wear, unfavorable exhaust-gas composition, higher fuel consumption, and unreliable vehicle heating. The temperature of the coolant must be regulated in order to counteract these unwanted side effects and to keep the coolant and engine temperature as constant as possible.

Expansion-element-regulated thermostat

A temperature-sensitive thermostat incorporating an expansion element proves to be a robust regulator operating independently of the changing pressure conditions in the cooling system. It is very similar in design to the map-controlled thermostat (Figure 6), but does not have a heating resistor. The expansion element used in this type of thermostat actuates a double-acting disk valve (main valve), which – until the operating temperature is reached – closes the connection to the radiator while simultaneously releasing coolant flow from the engine outlet to the bypass line (see Figure 2). The coolant flows uncooled back into the engine ("secondary cooling circuit").

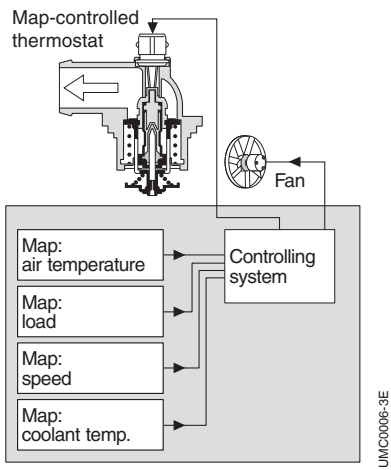
Both sides of the double-acting disk valve are partially opened within the control range of the thermostat. This allows a mixture of cooled and uncooled coolant to flow to the engine at such a rate as to maintain a constant operating tempera-

ture. At wide-open throttle (full load) the opening to the radiator is fully opened and the bypass line closed off ("primary cooling circuit").

Electronic map-controlled thermostat
Further possibilities to regulate the coolant temperature are permitted when a map-controlled thermostat is used. An electronically controlled thermostat differs from purely expansion-element-regulated thermostats in that the opening temperature can be controlled. A map-controlled thermostat features a heating resistor that is used to additionally heat the expansion element (Figure 6). This has the purpose of enlarge the opening of the double-acting disk valve to the radiator, thereby reducing the coolant temperature. The heating resistor is controlled by the engine management system in order to ensure that the engine operating temperature is optimally adapted to the operating conditions. The information necessary for this purpose is stored in the form of program maps in the engine-management system (Figure 7).

Raising the operating temperature in the part-load range and reducing the operating temperature at wide-open throttle provide the following benefits:

Figure 7: Electronic control of coolant temperature



– Lower fuel consumption,
– Low-pollutant exhaust-gas composition,
– Reduce engine wear,
– Improved heating efficiency of the vehicle interior.

Radiator design
The size and therefore the cooling capacity of a specific radiator can be determined by calculation, based on correlation equations derived from tests relating to thermal transfer and flow-pressure loss. The mass of air that flows through the radiator in the vehicle is a decisive factor and depends on the driving speed, the resistance to flow in the engine compartment, the resistance to flow of the radiator, and the efficiency of the fan.

The primary objective of radiator design is to maintain the coolant temperature at the engine outlet below a maximum permissible value under given operating conditions. Because at low driving speeds the dynamic pressure of the cooling air and thus the cooling-air flow rate in the radiator are very low, the cooling-air flow required for adequate cooling must be ensured either by a high-output fan or by a radiator with low flow resistance, while a high air-mass flow rate facilitates the use of smaller radiators with high flow resistance. However, the latter causes high energy consumption if the air-mass flow is generated by a powerful fan.

The task of determining the most favorable solution in terms of technical feasibility and economic efficiency is a matter of optimization that is best solved by the application of simulation tools. The most suitable and effective simulation tools describe all components that have an influence on air-mass flow. They depict the radiator as an integral heat-transfer medium or heat exchanger. The simulation results are checked by way of on-vehicle tests conducted in wind tunnels.

Intercooling (charge-air cooling)

Trends in engine development show a constant increase in specific engine output power. This development goes hand in hand with the current transition from naturally aspirated engines to supercharged engines and ultimately to supercharged, intercooled engines. The need for intercooling (charge-air cooling) is attributed to the higher air density levels associated with supercharging systems and therefore the amount of oxygen available in the combustion air. Intercooling also reduces exhaust-gas emissions from supercharged diesel engines. If there was no intercooling (charge-air cooling) employed on supercharged spark-ignition (SI) engines, appropriate steps would need to be taken to prevent engine knocking attributed to mixture enrichment or retarded ignition timing. Consequently, intercooling indirectly serves to reduce fuel consumption and exhaust emissions.

Design variations

Basically speaking, both the ambient air and the engine coolant can be employed to cool the charge air. With only few exceptions, air-cooled intercoolers are now used both on passenger cars and commercial vehicles.

Air-cooled intercoolers

Air-cooled intercoolers can be mounted in front of or next to the engine radiator, or even completely separate from the adiator at a different location in the engine compartment. A separately positioned intercooler can utilize either the unassisted vehicle air stream or its own fan. With the intercooler located in front of the engine radiator, the cooling-air fan ensures sufficient air flow even at low vehicle speeds. However, a drawback of this arrangement is that the cooling air is itself heated in the process. To compensate for this effect, the higher temperature of the inflowing cooling air must be taken into account in the design of the engine radiator.

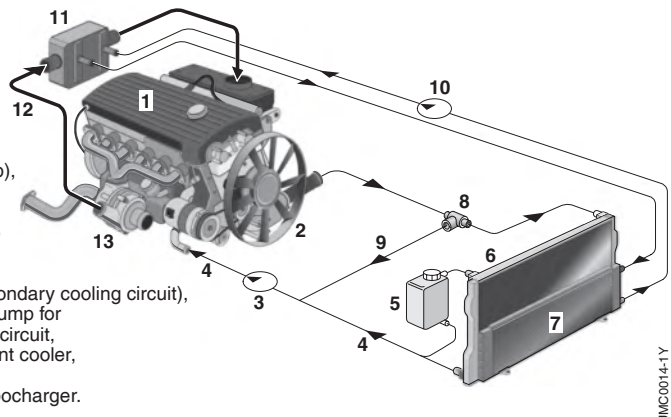
The system of corrugated aluminum fins and tubes employed for the intercooler core is similar to the design of the engine coolant radiator. Wide tubes with internal fins provide superior performance and structural integrity in actual practice. The fin density on the cooling-air side is relatively low and corresponds approximately to the density of the inner fins in order to achieve a good distribution of thermal-transfer resistance.

Wherever possible, the plenum chamber is injection-molded in fiberglass-reinforced polyamide as a single casting incorporating all connections and mounts. The plenum chambers which are subject

Figure 8: Passenger-car cooling system with intercooling

Engine cooling and indirect intercooling with separate low-temperature circuit.

1 Engine,
2 Fan,
3 Coolant pump,
4 Coolant line (arrow indicates direction of flow),
5 Expansion tank with pressure-relief valve (integrated in cap),
6 Main coolant radiator,
7 Low-temperature coolant radiator,
8 Thermostat,
9 Bypass line (secondary cooling circuit),
10 Auxiliary water pump for low-temperature circuit,
11 Charge-air coolant cooler,
12 Charge-air line,
13 Exhaust-gas turbocharger.



to increased stresses on the charge-air inlet side are injection-molded from highly heat-resistant PPA (polyphthalamide) or PPS (polyphenylene sulfide). They are flange-mounted to the radiator core and sealed by an integrated elastomer sealing element. Plenum chambers which feature undercut shapes or are intended for high-temperature applications are cast in aluminum and welded to the core.

Coolant-cooled intercoolers
Coolant-cooled intercoolers can be installed in virtually any location in the engine compartment, as there are no technical difficulties to supplying the system with coolant. In addition, thanks to its modest dimensions, this type of intercooler requires substantially less space than the air-cooled intercooler. Coolant-cooled intercoolers have a high power density. However, coolant at a very low temperature must be available in order to effectively cool the charge air. This requirement is of particular significance in commercial vehicles and heavy-duty trucks as, in this case, it is necessary to heat the charge air to a level of 15 K above the ambient temperature. Because this stipulation cannot be implemented with the temperature level of the normal cooling circuit at approx. 100 °C, it is necessary, in order to cool the charge air, to install a low-temperature radiator to ensure coolant is available in a separate circuit at the required temperature level.

Diffusion rate, intercooler
The diffusion rate Φ of an intercooler is particularly important in assessing intercooler performance. It defines the relationship between charge-air cooling efficiency and the charge-air/cooling-air temperature differential:

$$\Phi = \frac{T_{1E} - T_{1A}}{T_{1E} - T_{2E}}.$$

The equation's elements are:
T_{1E} Charge-air inlet temperature
T_{1A} Charge-air outlet temperature
T_{2E} Cooling-air or coolant inlet temperature

For passenger cars: Φ = 0.4...0.7
For commercial vehicles: Φ = 0.9 – 0.95

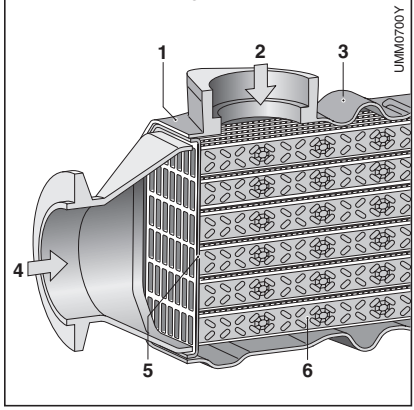
Exhaust-gas cooling

Exhaust-gas recirculation is used in diesel engines to reduce raw nitrogen-oxide emissions (NO_x). In gasoline engines it serves both to protect components and to dethrottle in the part-load range by charge dilution (see Exhaust-gas recirculation). Exhaust-gas recirculation delivers the best results when the recirculated exhaust gas is cooled in an EGR cooler.

Design of an EGR cooler
A central aspect in the design of an EGR cooler is, as well as its output, its durability. The former is achieved either by specially shaped tubes (winglets) which generate a high level of turbulence in the tube or by internal fins in the tubes which increase the surface area for heat exchange. The essential characteristic here is the tendency of an EGR cooler towards fouling. This can be significantly influenced by a suitable design of the winglets or tube fins. Figure 9 shows such an EGR cooler in longitudinal section. The winglet tubes direct the hot exhaust gas and are themselves surrounded by coolant.

It must be borne in mind with regard to durability that the hot exhaust gas may

Figure 9: EGR cooler
1 Stainless-steel housing (manufactured by internal high-pressure forming),
2 Coolant inlet,
3 Length compensation,
4 Exhaust-gas inlet,
5 Stainless-steel cooler base,
6 Stainless-steel winglet tubes.



condense in the cooler. The condensate produced has a lower pH value and attacks the material like a strong acid. For this reason, the gas-directing parts of EGR coolers are frequently made of stainless steel.

When the EGR cooler is in operation the tubes get hotter than the housing. Suitable design measures must be used to compensate for the linear difference in the cooler.

The issue of vibration also plays an important role in view of the fact that the cooler is usually bolted directly on the engine. The EGR cooler must therefore be as rigid as possible in its design. The first natural frequency of an EGR cooler should at least be above the first natural frequency of the engine block. For this reason, the brackets of such EGR coolers are specifically very rigid and sometimes multiply ribbed in design.

The cooler housing is often made of stainless steel, which can be strengthened further by cold-working. The exhaust pipes connected to the cooler should be isolated at the engine end. This isolation is achieved by bellows or by a comparable structure.

Cooling circuit

The coolant comes primarily from the engine cooling circuit, with the EGR cooler being cooled by coolant flowing out of the engine. A variant of this arrangement is “two-stage exhaust-gas cooling”. Here the exhaust gas is cooled in a first stage as described above. There is also a second cooling stage, which is supplied by a cooling circuit independent from the engine (low-temperature circuit). Because the temperatures in this second cooling stage are much lower (approx. 10 to 20 K above ambient temperature), the exhaust gas can also be cooled significantly below the otherwise normal values, increasing the effect of reducing nitrogen oxide. The two cooling stages are either integrated together in a single housing or designed separately as two independent components.

Oil and fuel cooling

Oil coolers are often needed in motor vehicles to cool both the engine oil and transmission oil. They are used when the heat losses from the engine or transmission can no longer be dissipated via the surface of the oil pan or the transmission with the result that the permitted oil temperatures are exceeded.

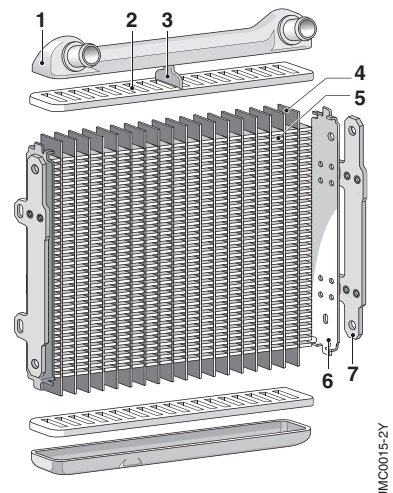
Air- or coolant-cooled oil coolers are used to suit their application profile.

Oil-to-air coolers

Oil-to-air coolers are predominantly made of aluminum. In most cases, soldered flat-tube/corrugated-fin systems (Figure 10, similar to engine radiators) with a high power density are used. Mechanically assembled systems with round tubes and fins are less commonly used. Turbulence inserts are brazed in the flat-tube system to increase cooling capacity and strength (to resist high internal pressures).

Figure 10: Oil-to-air cooler (flat-tube/corrugated-fin system)

1 Collector tank with oil connections,
2 Base,
3 Partition,
4 Flat tube with turbulence insert,
5 Corrugated fin,
6 Side section,
7 Bracket.

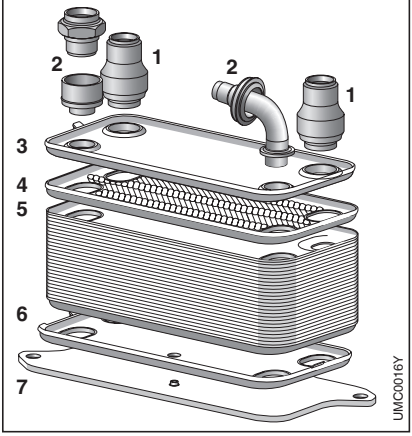


The principle of oil-to-air cooling is ideal for use in commercial vehicles and powerful passenger cars to cool the transmission oil. To ensure good ventilation, they are ideally mounted upstream of the coolant radiator, but can also be mounted elsewhere in the engine compartment. Oil-to-air coolers which are not located upstream of the main coolant radiator and are thus outside the range of action of the main fan must be supplied with cooling air accordingly, for example by exposure to dynamic air pressure or by a separate electric fan.

Oil-to-coolant coolers
Aluminum stack designs employed in oil-to-coolant coolers have largely replaced stainless-steel disk coolers, double-tube oil coolers, and aluminum forked-tube coolers.

Disk-stack oil coolers
Disk-stack oil coolers are made of individual disks with turbulence inserts inserted between the disks (Figure 11). The upright edges of the disks fit together in a casing. Passages connect the channels formed by the disks in such a way that coolant and oil flow through alternate channels.

Figure 10: Oil-to-coolant cooler (disk-stack oil cooler)
1 Oil connections,
2 Coolant connections,
3 Cover,
4 Disk stack (oil passage),
5 Disk stacks (coolant passage),
6 Reinforcement plate,
7 Base plate.



Double-tube oil coolers and flat-tube oil coolers
These cooler types are mounted directly in the radiator tank on the outlet side of the main radiator (coolant radiator). The radiator tank thus assumes the function of the coolant-side housing for the oil coolers. The double tube of the oil cooler is formed by an outer tube and an inner tube inside it with a turbulence insert in between. The two tubes are soldered to each other at the ends. The transmission oil passes through the space between the inner and outer tubes, around or through which already cooled coolant flows.

Double-tube coolers are used as transmission-oil coolers in passenger cars and commercial vehicle in the lower power segment up to approx. 2.5 kW. As power requirements increase up to 4 kW the double tube is replaced by numerous flat tubes connected to each other by turbulence plates on the coolant side. The flat tubes are interconnected by openings at their ends. The flat tubes are also provided on the oil side with turbulence plates, which are soldered to increase both cooling capacity and strength.

Disk oil coolers
Disk oil coolers are mounted between the engine block and the oil filter. They have a separate casing and a central channel for the oil to pass through. The oil flowing back from the oil filter is routed through a labyrinth of perforated disks separated by turbulence inserts. This labyrinth is cooled by coolant flowing from the main circuit through the casing.

Forked-tube coolers
Forked-tube coolers are made of finned forked tubes through which the coolant flows. On the oil side, they have no casing and must therefore be integrated into the oil-filter housing or the oil pan.

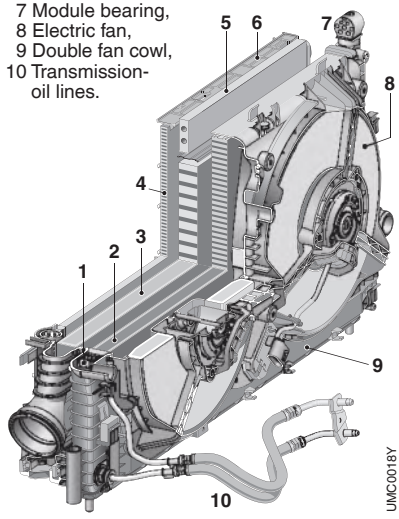
Engine-oil coolers for commercial-vehicle applications
The engine oil in commercial vehicles is generally cooled by stainless-steel disk-stack coolers or aluminum flat-tube cool-

ers without a casing on the coolant side. They are accommodated in an extended coolant duct in the engine block.

Fuel coolers
Fuel coolers are installed in modern diesel engines in order to cool excess diesel fuel down to permissible levels. This excess diesel fuel heats up during the injection process as the result of compression in the high-pressure pump, before it is routed back via the return line to the fuel tank. The excess diesel fuel that is returned without cooling to the tank has a temperature of well over 70 °C and heats the diesel fuel in the tank. If now the tank were almost empty, the temperature of the diesel fuel in the tank would therefore exceed the maximum permissible temperature. The excess fuel must therefore be cooled before being introduced into the tank.

The fuel can be cooled by means of an air-cooling or coolant-cooling system. Coolers operating along similar lines to oil-to-air coolers or disk-stack oil coolers are used to cool fuel.

Figure 12: Cooling module
1 Transmission-oil-to-coolant cooler,
2 Coolant radiator, 3 Intercooler,
4 Condenser,
5 Air-to-power-steering-fluid cooler,
6 Module frame,
7 Module bearing,
8 Electric fan,
9 Double fan cowl,
10 Transmission-oil lines.



Modularization

Cooling module
A cooling module is a structural unit which can consist of various components for cooling (e.g. main radiator, intercooler, oil-to-air cooler) and a condenser for air-conditioning a passenger car, and can include a fan unit with drive (e.g. electric motor) (Figure 12).

The design of a cooling module must take into consideration the interaction of the individual components, the dimensioning of the components with respect to the package space in the vehicle, and the procedure for dealing with interfaces. Important factors here are the mounting technology, the cooling-air ducts and the seals on the cooling-air side, the component connections on the fluid side, and the electrical plug-in connections.

The standard modular construction used in both passenger cars and commercial vehicles essentially offers a whole range of technical and economic benefits:
– Simplified logistics by combining components to form one structural unit,
– Reduced number of interfaces,
– Simplified mounting and assembly,
– Optimum component design by using matched components,
– Modular systems, encompassing various engine and equipment variants,
– Improved quality with regard to assembly.

Simulation and test methods are employed for the purpose of achieving optimized component design and layout in the cooling module. Assuming exact knowledge of the fan characteristics, the fan drive and the heat exchangers, simulation programs are created to replicate both the cooling-air side and the fluid side. By integrating the individual components into the simulation models, it is possible to examine the interactions of the individual components under various operating conditions. Ever greater significance is being attached to this type of virtual analysis, which is characterized by the use of computer-aided development tools. In line with this development, all geometric data are entered and processed in a CAD system (Computer-Aided Design). CFD analyses (Computational Fluid Dynamics) are con-

ducted to examine the flow of cooling air in the engine compartment, while FEM analyses (Finite-Element Method) provide statements concerning the strength and stability of the design layout. The design analysis phase concludes with verification tests that may also be performed in a wind tunnel and on vibration test rigs.

Cooling-system technology

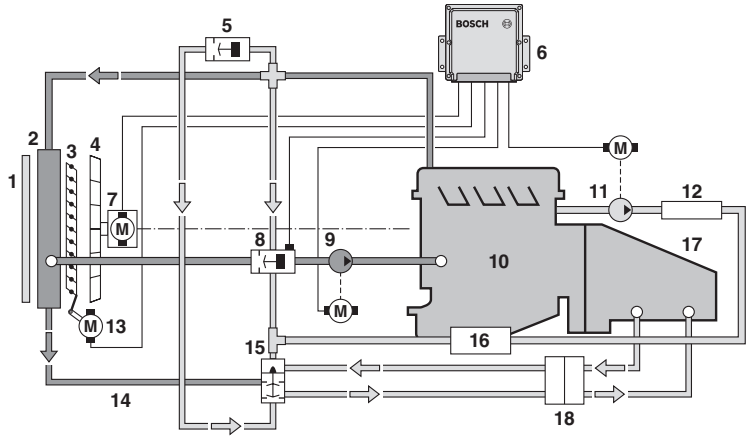
While the cooling module comprises a structural unit of components with defined functions, the cooling system encompasses all components that are associated with the functions of the cooling system, even if they do not form complete structural units (Figure 13). This includes for example, beyond the cooling module, lines, pumps, control components, and the expansion tank, if it is not already a component part of the cooling module.

Cooling-system technology, in which all the components are matched to each other, offers a whole range of technical and economic benefits:

– Reduced parasitic losses through appropriate hydraulic design matching,
– Consideration of control systems and dynamics,
– Consideration of the passenger-compartment heating system,
– Larger scope of intervention options for the purpose of optimizing the design,
– Standardized assembly concept for all cooling-system components,
– Reduced development expenditure by cutting the number of development interfaces.

Figure 13: Controlled cooling system

Example architecture with heat exchangers and actuators.
1 Condenser, 2 Coolant radiator, 3 Shutter, 4 Fan,
5 Thermostat, 6 ECU, 7 Electric motor (fan drive),
8 Electronically controlled thermostat,
9 Electric motor with pump, 10 Internal-combustion engine,
11 Electric motor with pump, 12 Heater core,
13 Stepper motor, 14 Low-temperature circuit,
15 Low-temperature regulator, 16 Oil cooler,
17 Transmission, 18 Transmission-oil cooler.



Intelligent thermal management

Function

Future trends are heading toward the system-optimized regulation of various heat and substance flows. Thermal management goes beyond cooling-system technology in that it takes into consideration all material and heat flow systems in the vehicle, i.e. in addition to the flow structures of the cooling system, it also deals with those of the air-conditioning system. The objectives of optimization consist in reducing fuel consumption and exhaust emissions, increasing air-conditioning comfort, increasing the service life of components, and improving cooling capacity in part-load states.

Optimization objectives

One of the basic principles of thermal management concerns itself with the fact that auxiliary energy employed to operate the cooling system always represents a loss for the vehicle energy balance, and component efficiency cannot be increased arbitrarily at a constant supply of auxiliary energy. To achieve the optimization objectives, the cooling system is therefore equipped with "intelligence", installed in familiar and new types of actuators, as well as in microprocessor-controlled control systems that operate these actuators. One example is demand-triggered regulation of cooling-air flow by radiator shutters and controllable fan drives so that cooling-air throughout is kept to a minimum under all operating conditions. In addition to improving the vehicle's drag coefficient (c_d), this measure also ensures that all media achieve their operating temperature more effectively during the warm-up phase after a cold start, and that the passenger compartment is heated more efficiently. Cutting back the use of auxiliary energy in this way means that auxiliary energy can be diverted for use in operating states that are critical to the cooling output, while still achieving the optimization objectives.

Another important basic principle is to maintain a constant temperature in the components to be cooled as far as possible, irrespective of the operating state and the ambient conditions. An example of this temperature control principle is to use coolant to regulate the transmission oil temperature. Heating the transmission oil during the warm-up phase and employing an efficient cooling system to prevent the transmission oil from overheating reduce friction losses in the transmission, increase the service life of the transmission, and extend the service intervals for the transmission oil.

Ultimately, considering the cooling and air-conditioning systems in their entirety opens up the option of utilizing "thermal integration". Heat flow from one of the systems can be utilized or dissipated by another system without the need for any major additional input in auxiliary energy. An example of this is the utilization of waste heat from the exhaust-gas cooling system to heat up the vehicle interior.

Applications

In the field of engine cooling thermal management is used for the following:

- Transmission-oil temperature equalization,
- Map-controlled thermostat,
- Electrically controlled viscous clutch,
- Controllable electric coolant pump,
- Cooling-air control, e.g. radiator shutter,
- Exhaust-gas cooling,
- Coolant-cooled charge-air cooling (intercooling).

Advantages

The fuel-saving potential based on the sum of all measures is in the range of approx. 5% (for passenger vehicles). On top of this, there is a range of further advantages corresponding to the above-mentioned optimization objectives. The extent to which engine management utilizes the cooling system control options is of decisive significance in leveraging this potential.

In the meantime, individual measures have been implemented to achieve system-optimized temperature equalization in motor vehicles. Nevertheless, thermal management as an all-encompassing optimization principle has to date not been fully implemented and remains the reserve of future vehicle generations.

Engine lubrication

Force-feed lubrication system

Heat and wear are produced in engines by the friction of moving parts. Small amounts of residual dirt from production and particulate ingress from the outside are also to be expected in the engine system. The engine system is lubricated with lube oil in order to minimize friction and eliminate or reduce wear and to remove particulates and heat from the engine system.

The force-feed lubrication system in combination with splash and oil mist lubrication is the most commonly used system for lubricating motor-vehicle engines. An oil pump (usually a gear pump) conveys pressurized oil to all lubricating and bearing points in the engine (Figure 1), while the sliding parts are supplied with lube oil and lubricated by splash lubrication systems and oil mist.

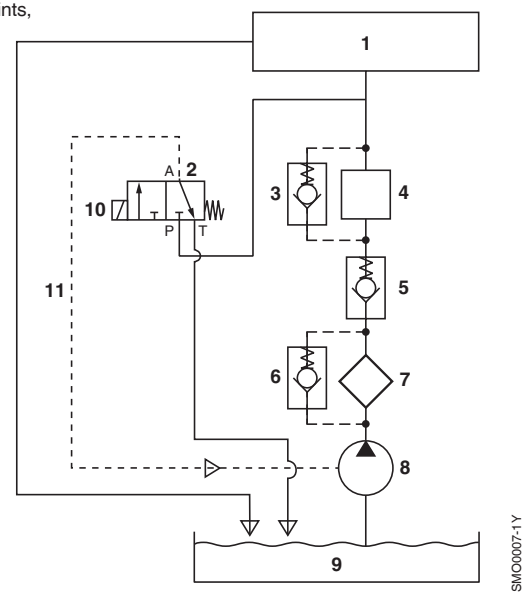
After flowing through the bearing surfaces and sliding parts, the oil collects below the engine in the oil pan. The oil pan is a reservoir where the oil cools, and the foam dissipates and settles. In the meantime an additional oil cooler is used in virtually all engines to regulate the oil temperature. In addition to filtration of the oil, thermal management is of particular importance.

Engine service life can be prolonged significantly by keeping the oil clean, i.e. filtration. Oil cooling by means of good thermal management makes possible current technologies such as single or multiple charge-air charging and the associated downsizing of engines.

Figure 1: Schematic diagram of a forced-feed lubrication system

(Schematic diagram).
1 Engine lubricating and bearing points,
2 Solenoid valve,
3 Cooler bypass valve,
4 Oil-to-water heat exchanger,
5 Anti-drainback valve,
6 Filter bypass valve,
7 Oil filter,
8 Variable oil pump,
9 Oil pan,
10 Signal from engine control unit,
11 Signal for pressure-stage preselection.

A Working line:
Pressure signal for oil pump.
P Pressure line:
Oil-gallery pressure.
T Ventilation line:
Ventilation to crankcase.



Components

Oil filters

Function

Oil filters remove and reduce particulates (soot and other residues from incomplete combustion, metal abrasion, dust, etc.) from the engine oil, which could otherwise cause damage or wear in the lubrication circuit. Wear in this context means that the particulates can get between the moving engine components, e.g. between piston and cylinder wall. Here the surfaces can be damaged by the particulates, which in turn can cause scoring or further particulate production. Since the engine oil is constantly circulating in the lubricating oil system, inadequate filtration could cause particulates to accumulate and this would accelerate the rate of wear of components that move against each other. The oil filter does not filter out liquid or soluble constituents such as water, additives or decay products attributed to the effects of oil aging.

The typical size of particles in engine oil ranges from 0.5 to 500 µm. The fineness of the oil filter is therefore specifically adapted to the requirements of a particular engine.

Different types and designs

In principle, oil filters are based on two specific designs: easy-change filters and housing filters. In an easy-change oil filter (Figure 2), the filter element is located in a housing that cannot be opened and which is secured by means of a threaded stud to the engine block. The complete easy-change filter unit is replaced as part of the oil service.

The housing oil filter (Figure 3) is made up of a housing that is permanently connected to the engine block and can be opened to access the replaceable filter element. During the oil service, only the filter element is replaced; the housing is a permanent component. The filter element as a rule is based on a metal-free design. This means that the filter element can be completely incinerated.

Figure 2: Easy-change oil filter
1 Threaded cap,
2 Filter element,
3 Filter bypass valve,
4 Spring,
5 Seal,
6 Anti-drainback diaphragm,
7 Housing
8 Center tube.

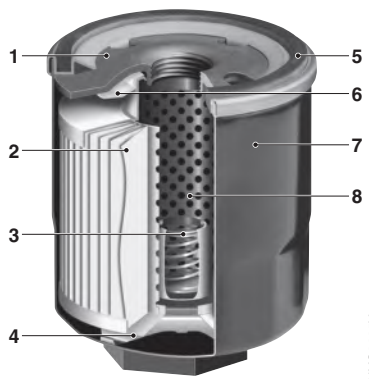
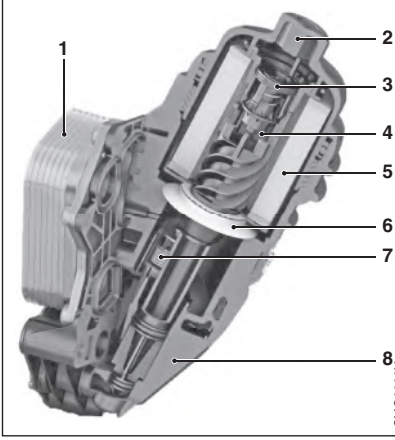


Figure 3: Housing oil filter
1 Oil-to-water heat exchanger,
2 Screw cap,
3 Filter bypass valve,
4 Center tube,
5 Filter element,
6 Anti-drainback diaphragm,
7 Anti-drainback valve,
8 Filter housing.



Besides the filter element, the two filter designs normally feature a filter bypass valve, which opens at high differential pressures to ensure effective lubrication at the necessary points in the engine. Typical opening pressures range from 0.8 to 2.5 bar. Elevated differential pressures can occur in connection with high oil viscosities or when the filter element is already heavily fouled.

Depending on the specific engine requirements, the two filter designs may also feature a non-return or backflow check valve on the filtered or unfiltered oil side (contaminated oil side). These valves prevent the oil filter housing from draining empty after the engine is turned off.

Currently, the oil and oil-filter change intervals for passenger-car engines are between 15,000 and 50,000 km, and for commercial vehicles are between 60,000 and 120,000 km – very much dependent on the engine system, place of use and the service strategy of the car manufacturer.

Filter media

Various types of deep-bed filter media are employed in oil filtration. These are entangled-fiber structures (nonwovens) which exist in different configurations. The filter material most commonly used is flat media which, in the majority of cases, is pleated, but in some applications they are also wound or used in the form of fiber packings, especially in bypass oil filters. A mixture of cellulose with synthetic fibers is the most widely used fiber material. The synthetic-fiber content can vary markedly depending on the requirements. Alternatively, amounts of glass fibers can be mixed with the cellulose. These filter media are impregnated with resin to enhance their resistance to oil and their operational stability.

However, filter media made of purely synthetic fibers are being used even more, as they have a much greater resistance to chemicals and clear benefits in terms of differential-pressure behavior. These attributes allow for longer service intervals. They also provide better options for structuring the three-dimensional fiber matrix to specifically optimize filtration and increase particulate retention efficiency.

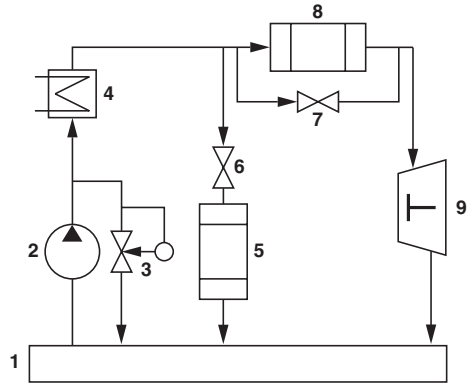
Full-flow oil filters

All vehicles are as a rule equipped with full-flow oil filters. Based on this filtration principle, the entire volumetric flow of oil that is pumped to the lubricating points in the engine is routed through the filter (Figure 4). Consequently, all particles which could cause damage and wear due to their size are trapped the first time they pass through the filter.

The decisive factors governing the filter area are oil volumetric flow and particulate retention capacity.

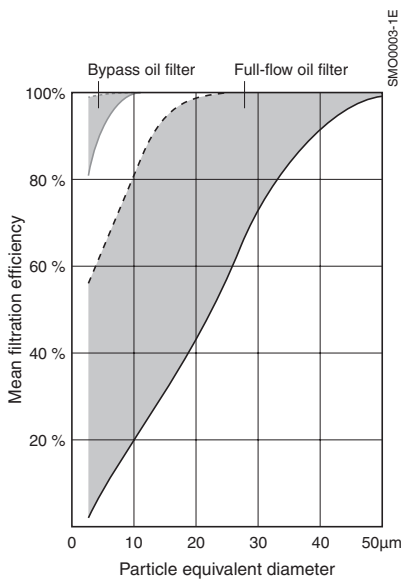
Figure 4: Oil circuit with full-flow and bypass oil filters

(Schematic diagram).
1 Oil pan,
2 Oil pump,
3 Pressure-control valve,
4 Oil cooler,
5 Bypass oil filter,
6 Throttle,
7 Bypass valve,
8 Full-flow oil filter,
9 Engine.



Bypass oil filters
Bypass oil filters, designed as deep-bed filters or as centrifuges, are used for superfine or microfiltration of the engine oil. These filters remove much finer particles from the oil than is possible using full-flow oil filters (Figure 5). They can remove minute abrasive particles to enhance wear protection. Soot particles are also filtered out to reduce any increase in oil viscosity. The maximum permissible soot concentration is roughly 3 to 5 %. Oil viscosity increases substantially at higher soot concentrations, resulting in a drop in the operational effectiveness of the oil. For this reason, bypass filters are used mainly on diesel engines. Only part of the oil flow (8 to 10 %) from the engine is routed via the bypass oil filter.

Figure 5: Filter fineness of full-flow and bypass oil filters (acc. to ISO 4548-12 [1])
------- Maximum filtration efficiency for bypass or full-flow oil filter.
——— Minimum filtration efficiency for bypass or full-flow oil filter.



Thermal management

Function
Aside from its lubricating function, the engine oil also has the task of transferring or carrying off heat. At the same time the engine-oil temperature must not exceed a critical value, as otherwise there is a risk of thermal destruction, so-called cracking of the oil.

Oil-to-water heat exchangers, which are frequently combined with the oil filter module, are used to control the temperature of the oil circuit. This combination allows for a more compact construction and the reduction of interfaces. The primary task is to transfer heat from the oil circuit to the water circuit. The heat energy is then dissipated via this to atmosphere for example with an air-to-water heat exchanger.

The heat transfer must be controlled to provide optimum thermal management. To quickly heat the engine oil after cold starting, the oil-to-water heat exchanger should only be activated after an increase in temperature in the oil. This can be done by thermostatic valves which are controlled with wax elements or thermo-springs. Electric solenoid valves provide for even better control options.

Oil-to-water heat exchanger
The preferred design for an oil-to-water heat exchanger is with plates stacked on top of one another. Oil and water flow alternately through the chambers. The heat is transferred via the plates to the neighboring fluid. To achieve high heat-transfer rates the flow in the columns should be as turbulent as possible. Structures are introduced into the heat-transfer plates or turbulence plates installed in the column for this purpose.

References
[1] ISO 4548-12: Methods of test for full-flow lubricating oil filters for internal combustion engines – Part 12: Filtration efficiency using particle counting and contaminant retention capacity.

Air-intake systems and intake manifolds

Overview

Functions

The air-intake system and the intake manifold are designed to supply the engine with air that is as cold and particle-free as possible, and to distribute this air to the individual cylinders. The main functions of the air-intake system are to direct the air from the front of the vehicle to the engine, to separate particulates contained in the intake air from the surrounding area, and to muffle the noises emitted by the engine.

Components

Basic functions

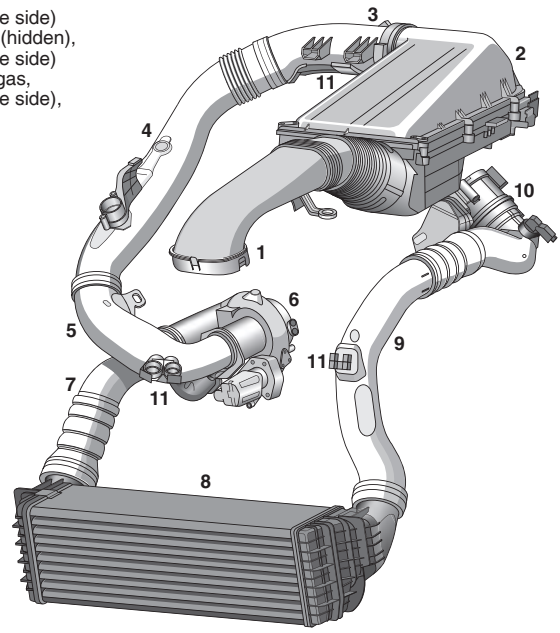
The air-intake system (Figure 1) consists of an air duct on the raw-air side, the air filter as the central components, and the lines to which the throttle valve and the intake manifold are connected. The intake manifold distributes the intake air to the individual cylinders and ensures uniform and effective combustion. The air filter prevents mineral dust and particles from being drawn into the engine and contaminating the engine oil. This reduces wear in the bearings, piston rings, cylinder walls, etc. In this way the individual system components can have a crucial positive effect on the performance and service life of the engine.

Supplementary functions

The air-intake system often incorporates further function components such as, for example, sensors (air-mass meter, pressure sensor, and temperature sensor), HC adsorber, water separator, and even anti-snow systems.

Figure 1: Design of an air-intake system with exhaust-gas turbocharging

1 Raw-air line,
2 Air-filter housing,
3 Clean-air line 1 (low-pressure side) with hot-film air-mass meter (hidden),
4 Clean-air line 2 (low-pressure side) with introduction of blow-by gas,
5 Clean-air line 3 (low-pressure side),
6 Exhaust-gas turbocharger,
7 Charge-air line (hot high-pressure side),
8 Intercooler,
9 Charge-air line (cold high-pressure side),
10 Throttle valve,
11 Attachment points.



Requirements

Functional requirements

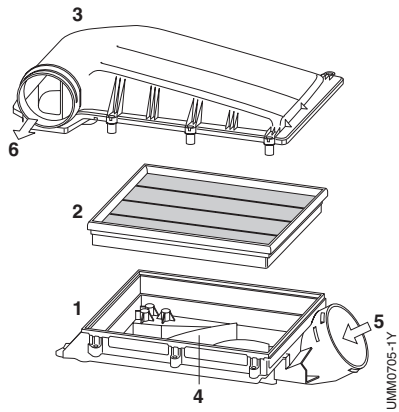
The development of the air-intake system and its components is based on numerous specifications which are shaped on the one hand by space considerations and on the other hand by functional requirements. Specifications define for example temperatures in the engine compartment which influence the choice of plastic materials and elastomers for the individual components, permissible leakage rates and pressure losses for the system or the individual components, the acoustic performance of the intake tract, and the separation performance with regard to filtration as service life and as filtration-efficiency requirement.

With regard to the intake manifold, further emphasis is placed on pressure-pulsation resistance and the level of uniform distribution of intake air. The air flow into the combustion chamber can if necessary be positively influenced by different types of flap.

Visual appearance

The visual appearance and the design of surfaces in the engine compartment gain in importance as soon as surfaces are located in the visible area of the engine compartment (e.g. design air filters or engine-mounted air filters).

Figure 2: Design of an air filter
1 Raw-air housing, 2 Air-filter element,
3 Clean-air hood, 4 Flow deflectors,
5 Raw-air connection, 6 Clean-air connection.



Passenger-car air-intake system

Air duct

The intake point of the raw-air line is usually located at the front of the vehicle. It is particularly important in that it has a crucial influence on the ingress of particles and the unwanted ingress of water droplets when it is raining and of snow crystals when it is snowing. However, this position is generally stipulated by the automobile manufacturers.

Classic internal-combustion engines are divided into non-supercharged naturally aspirated engines and supercharged engines, in which an exhaust-gas turbocharger for example is deployed. The air-intake system (Figure 1) in both variants consists of a raw-air line, which feeds into an air-filter housing with integrated filter element, and a clean-air line on the low-pressure side, which in a naturally aspirated engine serves as the interface to the throttle valve. In a turbocharged engine the clean-air line on the low-pressure side feeds into the turbocharger, while the charge-air lines on the hot and cold sides lead to the intercooler and to the throttle valve respectively. Low pressure losses can be achieved by large line cross-sections (this influences space and acoustics) and by optimized flow routing.

The lines are usually made of plastic and elastomer materials, while the air-filter housings are generally injected-molded plastic parts. Depending on the application, the air filter can be mounted to the body and to the engine in such a way that the lines compensate not only the tolerances between the components and their attachment points, but also above all the movement of the engine relative to the body. The components are joined with vibration isolation to the vehicle by way of rubber buffers.

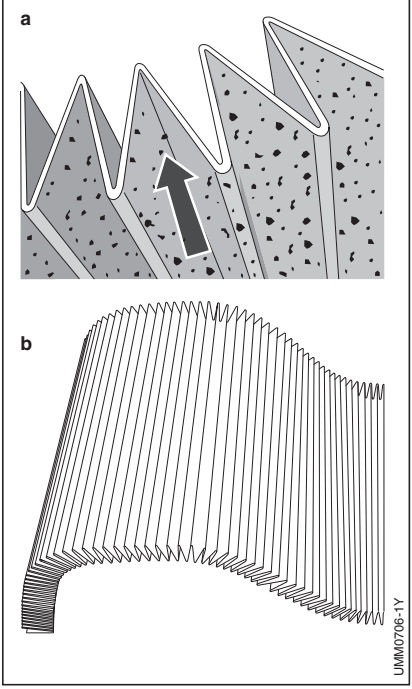
Air filtration

Engine air-intake filters (Figure 2) reduce the particles contained in the intake air which can be generated for example by abrasion and incomplete combustion or condensation processes or are even of natural origin in the form of organic and mineral dust.

Composition of air pollutants
Typical air pollutants includes oil mist, aerosols, diesel soot, industrial waste gases, pollen, and dust. These particles vary greatly in size. The dust particles drawn in together with the intake air have a diameter of between 0.01 µm (mostly soot particles) and 2 mm (sand grains). Approximately 75 % of the particles (referred to the mass (weight)) are between 5 µm and 100 µm.

The amount of dust contained in the intake air depends heavily on the environment in which the vehicle is used (e. g. freeway/interstate or dirt track). In extreme cases, over a period of ten years, the mass concentration accumulated in a passenger car can range between a few grams through to several kilograms of dust.

Figure 3: Filter media
a) Flow through a folded/pleated filter medium,
b) Folded/pleated filter medium.



Filter elements
Filter elements which incorporate the latest technology achieve total mass filtration rates of up to 99.8 %. It is necessary to maintain these values under all prevailing conditions – also under dynamic conditions as found in the engine's intake tract. Poor-quality filters will exhibit an increased rate of dust breakdown or rupture.

Design
Filter elements are designed to meet the requirements of each individual type of engine. This ensures that pressure losses remain minimal and that high levels of filtration efficiency are not dependent on air throughput. The filter medium that makes up the filter elements in flat or cylindrical filters is installed in folded or pleated layers in order to achieve a maximum filter surface area in the smallest possible space (Figure 3). Conical, oval, stepped, as well as trapezoidal geometries complement the standard structures with the aim of optimizing the use of package space that is becoming very scarce and confined in the engine compartment. These media, mostly consisting of cellulose fibers, go through a special embossing and impregnation process to achieve the necessary mechanical and thermal strength, sufficient water stability, and resistance to chemicals. Newer applications are equipped with flame-inhibiting filter media to reduce the risk of fire caused for example by an drawn-in cigarette.

The elements are designed in accordance with the spacings and replacement intervals defined by the vehicle manufacturer. The intervals are between two and four, and sometimes even six years, or 30,000 km up to 100,000 km.

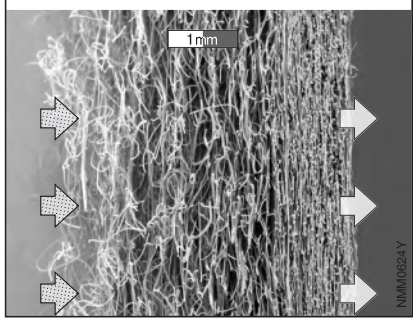
Filter media and design
The media used in passenger-car air filters are mostly part of deep-bed filters which, in contrast to surface filters, retain the particles in the structure of the filter medium. Deep-bed filters with a high dust absorption capacity are always used to advantage wherever an economic solution is required to filter large volumetric flow rates with low particulate concentrations.

The demands for compact, high-performance filter elements (with reduced package space), as well as longer service intervals, are the driving force behind the development of new, innovative air filter media. To further increase the dust retention capacity of deep-bed filters, design engineers are increasingly using media which have a gradient structure with increasing fiber density to the clean-air side.

Fully synthetic filter media
New air-filter media consisting of synthetic fibers with much improved performance data have already been phased into series production. Figure 4 shows a photograph of a synthetic high-performance filter medium (felt) with continuously increasing density and decreasing fiber diameter across the filter section from the input side to the output side.

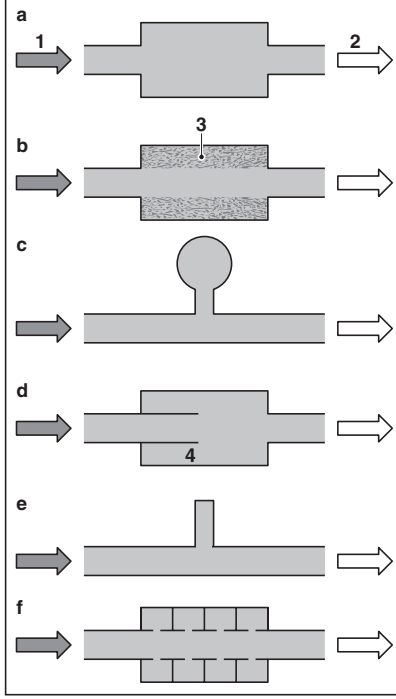
Semi-synthetic filter media
Better results than with purely cellulose-based media can be achieved using composite materials. Here, for example, a cellulose-based paper layer is combined with a synthetic layer of melt-blown fibers. The melt-blown fibers constitute a layer of fibers formed from a polymer melt in the air flow which is either deposited directly on the paper layer or laminated on in a separate process step as a separate layer.

Figure 4: Photographic image of an air-filter medium consisting of synthetic fibers, taken with a scanning electron microscope



Acoustics
The air-intake system contributes in a similar way to the exhaust system to the overall noise of the vehicle. It is therefore necessary to introduce noise-reduction measures already in the early development stages. In the past these measures

Figure 5: Acoustic muffling measures with different resonator shapes
a) Resonator-type muffler (relatively wide-band, suitable for low and medium frequencies),
b) Absorption-type muffler (wide-band, suitable for medium and higher frequencies),
c) Resonance muffler (narrow-band, suitable for low and medium frequencies),
d) Whistle muffler (narrow-band, high muffling, suitable for medium frequencies above the Helmholtz resonance),
e) Branch whistle muffler ($\lambda/4$-pipe, narrow-band, suitable for medium frequencies),
f) Wide-band muffler for turbocharged engines (wide-band, suitable for higher frequencies).
1 Air duct, upstream,
2 Air duct, downstream,
3 Acoustically absorbent material,
4 Inserted pipe.



were confined to designing a sufficiently sized air-filter housing as an intake muffler. In the meantime the excitation spectrum has shifted from low frequencies (f < 200 Hz) to medium to high frequencies with the result that separate measures in the form of Helmholtz resonators and $\lambda/4$ pipes are used. Computer simulation programs that determine the optimum number of resonators and their frequency are used to design these measures.

Wide-band mufflers for turbocharged engines

The increasingly widespread use of turbocharged engines has led design engineers to focus their attention on new noise sources. The whistling and whining noises are subjectively perceived to be disruptive and annoying and now measures are often integrated in the form of wide-band mufflers in the clean-air line of the air-filter system. Wide-band mufflers comprise several individual chambers through which a perforated central pipe passes (Figure 5). The individual resonant frequency of a chamber is determined by the chamber volume and the degree of perforation; a wide-band muffling effect is achieved by arranging several chambers in succession. Here too modern simulation programs are used to optimize the number.

In addition to resonator-type mufflers, absorption-type mufflers as featured in exhaust systems are occasionally used.

HC adsorption

For the Californian market and some other US states limits regarding the emission of highly volatile hydrocarbons are defined for vehicles (e.g. SULEV, Super Ultra Low Emission Vehicle). The US environment agency EPA is responsible for implementing the legal requirements regarding HC emission limits (High Volatile Hydro Carbons) within the framework of a multiple-stage regulation. The engine manufacturers must comply with these statutory provisions and adhere to time limits after which engines must feature pollutant-emission-reducing technologies to satisfy the emission standards.

Some of the unburned hydrocarbons (primarily fuel components) escape after the engine is turned off from the engine through the intake manifold and from the crankcase ventilation into the air-intake system, travel against the normal direction of air flow, and exit the system to atmosphere. This can mean that it is necessary to integrate an HC adsorber on the clean-air side of the system so that the hydrocarbons are separated by adsorption on activated-carbon media. An equivalent, the "Butane Working Capacity" (BWC), is defined in specifications in order to define the absorption capacity comparably. The adsorber elements can be used in the "full flow" (main path, full throughflow) or in a bypass configuration.

Unlike air-filter elements, adsorber elements according to the requirements are lifetime components which may not be removed from the system. The elements are charged while the engine is stopped with the air-borne hydrocarbons of minimal droplet size, and desorption takes place the next time the engine is operated, i.e. the hydrocarbons are returned to the combustion process in this cycle and the HC adsorber is regenerated for the next charging cycle. The adsorptive medium used is activated carbon, which is located as a single or double layer between the carrier media. These are for the most part elements which are manufactured in a process as plastic coating.

Water separation

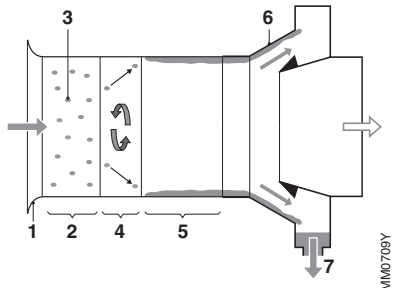
When driving in the rain, not inconsiderable amounts of water may be drawn in, depending on the position and layout of the intake fitting at the front of the vehicle. Because the filter medium only has a limited water absorption capacity, the drawn-in liquid may reach the clean-air side of the system after a certain time. Signal deviations or even damage to the hot-film air-mass meter may occur if droplets touch the hot sensor surface and result in a local, unwanted cooling of this surface. The signal deviation suggests an incorrect air-mass flow and gives rise to an incorrect setting of the mixture composition, to power losses, and ultimately to increased consumption (and increased

CO_2 emissions). The time until water penetration at the filter element can be significantly extended or even avoided entirely by way of upstream measures.

Finned separators, baffle plates or measures based on the cyclone principle such as the "peeling collar" (Figure 6) integrated in the intake air are used to separate water droplets. The shorter the raw-air line – i.e. the distance from the air inlet to the filter element – the more difficult the task. Essentially, the aim is to achieve additional flow pressure losses that are as low as possible. The measures must be combined with suitable water-discharge devices which divert the separated water from the system to atmosphere. However, separating principles are also used which carry the liquid in combination with measures on the housing side and with additional water-separating fleeces attached to the filter element from the system.

An optimized design of the raw-air line and the air-filter housing clearly enhance the above-mentioned measures such that outstanding separation results are achieved even with comparatively short raw-air lines.

Figure 6: Principle of water separation
1 Raw-air intake,
2 Raw-air line,
3 Water droplets,
4 Swirl generator (optional),
5 Wall-applied water film,
6 Peeling collar,
7 Water discharge.

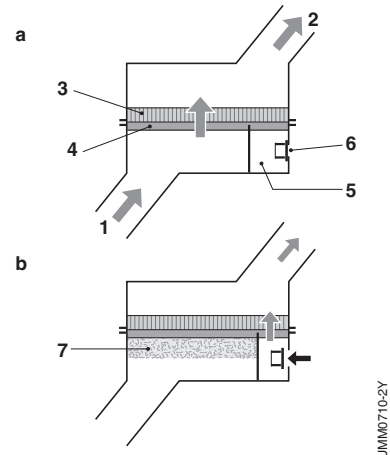


Anti-snow system

In winter when it is snowing snow crystals may be drawn into the system which charge the filter element on the raw-air side. If this situation persists, the initial upshot will be an increase in pressure loss. In due course the filter element may become blocked on the raw-air side such that the engine stops receiving sufficient intake air for the combustion process. This will ultimately cause the engine to stall.

To prevent this, anti-snow systems (ASS) – essentially a second intake – are used (Figure 7). If the filter element becomes blocked, hot, dry and snow-free air is drawn in from the engine compartment, whereby this reduced air flow maintains the engine function. The supplementary intake can be opened by temperature- or pressure-controlled control mechanisms. In the case of thermal control a flap is moved by an expanding-wax element. Pressure control is based on a spring-loaded valve integrated in the

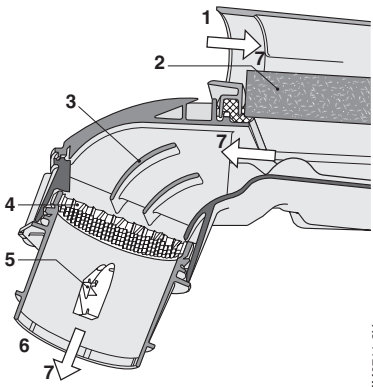
Figure 7: Anti-snow system
a) Bypass valve closed,
b) Bypass valve open.
1 Raw air,
2 Clean air,
3 Air-filter element,
4 Fleece on raw-air side,
5 Dry chamber,
6 Bypass valve,
7 Snow block.



air filter on the raw-air side. The filter element is basically fitted with a preliminary fleece which assumes the sealing function and protects the “dry” chamber of the second intake against snow ingress.

Sensor technology
Sensors which measure the pressure, the temperature and even the air-mass flow are often integrated on the clean-air side of the air filter. The exhaust-gas-relevant air-mass flow is measured by a hot-film air-mass meter. As the emission-control legislation is tightened the requirements with regarding to signal quality are becoming increasingly stringent. Signal deviations for an air-filter system in the range of ±2.5% are difficult to achieve, but not unusual. This requires an increased simulation outlay to identify disruptive turbulent areas and instabilities in the flow close to the sensor and to minimize them by means of suitable product design and additional measures or components (e.g. air-duct grills, fins or air deflectors) (Figure 8).

Figure 8: Flow optimization upstream of the hot-film air-mass meter
1 Raw-air connection,
2 Air-filter element,
3 Deflector geometry,
4 Rectifying grill,
5 Hot-film air-mass meter,
6 Clean-air connection,
7 Flow direction.



Visual design of surfaces
The structuring of plastic surfaces in the visible area called for by customers for visual reasons can be achieved in a variety of ways, such as for example by erosion, blasting or photoetching of the injection-molding die. Engine-mounted design air filters are a special example here.

Interfaces
The type of connection point between the air-filter housing and the clean-air line is often geared towards customer requirements. The production plant of the automobile manufacturer often specifies the shape of the interface. If an elastomer component is called for, lines based on the material PP-EPDM with integrated bellows for isolation are manufactured by injection molding. These can be connected to a line section made using blown-part technology.

Crankcase ventilation
The clean-air line on the low-pressure side can incorporate further interfaces such as for example the blow-by introduction point from the crankcase ventilation or even the extraction point for the PCV (Positive Crankcase Ventilation). The blow-by introduction point is frequently fitted with a heating tube which is intended to prevent icing.

Low-pressure exhaust-gas recirculation
A measure specially suited to smaller diesel engines to meet the strict Euro 6 limits for nitrogen oxides (NO_x) is low-pressure exhaust-gas recirculation (see Exhaust-gas recirculation, EGR). But even larger-sized engines utilize low-pressure exhaust-gas recirculation, often in conjunction with SCR systems (Selective Catalytic Reduction). With low-pressure exhaust-gas recirculation the exhaust gas is diverted after the diesel particulate filter, cooled in an EGR cooler, and introduced into the clean-air line upstream of the turbocharger compressor. To obtain high EGR rates in wide ranges of the engine map, it is very important for the EGR system to have a low-pressure-loss

design including mixing of air with exhaust gas. For special operating points in the part-load range an additional pressure drop is generated in the EGR system by an exhaust-gas restrictor (downstream of the EGR extraction point) or by an air-intake restrictor in the clean-air line (upstream of the EGR introduction point). Configuring the mixer geometry appropriately prevents local temperature gradients from damaging the compressor and compressor efficiency from being reduced. After the EGR cooler the exhaust gas reaches maximum temperatures of 150 to 200 °C, which enable the mixing and control functions to be executed cost-effectively and space-efficiently with the aid of plastic components. However, it is important when choosing the material to ensure that it is resistant to acidic condensate (pH value of 2 to 3) which can be produced at exhaust-gas temperatures below 85 °C.

The use of low-pressure exhaust-gas recirculation is being analyzed even in gasoline engines; however, the objectives here are to reduce CO_2 and the knock tendency rather than to reduce NO_x emissions. The first applications in series production are expected in the next few years.

Validation of air filter

Full product validation, consisting of simulation, acoustics and component testing, is an essential part of the development process for air-filter systems in order to achieve a high degree of product maturity as early as possible. Already in the early concept phase – within the framework of virtual product validation – simulations are conducted to optimize the functions and properties of air-filter systems. Here the acoustic properties such as intake-opening noise and surface sound radiation are calculated and optimized.

As well as acoustic behavior the fluid properties such as pressure loss and approaching flow to the measuring-sensor system are analyzed and specifically improved. In addition, the vibration characteristics are tested under roadway or engine excitation using finite-element simulations to fulfill the later life test without any problems. The filtration properties among others are checked as soon as the first sample components are available. The scope of testing within the framework of component testing includes tests on individual components and on the complete system. This is rounded off by functional tests where the ambient conditions prevailing in the vehicle (such as temperature, moisture, and media pressurization) are superimposed.

The component life is tested in comprehensive vibration and pulsation tests which simulate the real load in the vehicle. In view of the strict emission-control legislation, the flow behavior in the air filter must also be very closely analyzed with a dust charge also taken into consideration. Influencing factors such as fold/pleat geometry of the air-filter element and spread of the filter medium are taken into consideration here. Component validation is rounded off by simulations of the injection-molding process to optimize component quality and the manufacturing process.

Trends – further development of air-intake systems

Three main trends can currently be observed in the market: Downsizing of engines, platform strategies to utilize synergy effects, and stricter requirements with regard to emission-control legislation.

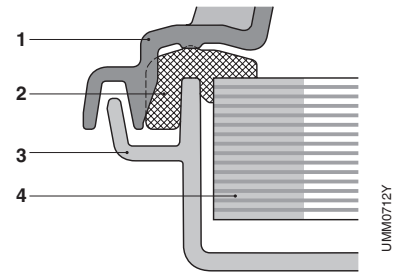
Downsizing necessitates a higher proportion of supercharged engines with the following effects: Specifically higher air-mass flows, smaller and more complex air-filter housings, increased conflict between acoustics and pressure loss, higher temperatures in the engine compartment, and higher charge-air pressures. This necessitates a higher simulation outlay for example to optimize flow and acoustics, new air-filter media, innovative filter-element designs, and higher-quality plastic materials.

Platform strategies of the automobile manufacturers necessitate more complex package spaces and a high-grade identical-part strategy. Stricter requirements

with regard to emission-control legislation give rise to heightened specifications with regard to sensor technology, particularly the quality of the signal stability of the hot-film air-mass meter. This likewise necessitates more complex simulation tools, new test procedures, and more complex product designs and constructions.

The weights of components can be reduced with new production technologies. Air filters produced by injection molding on the basis of foamed plastics help to achieve this. A further possibility is to reduce wall thicknesses, where the acoustic behavior in particular of the air-filter housing must be observed. Structure-borne noise radiation can be countercompensated by cambering (curved geometry) surfaces and integrating ribs or beads (groove-shaped depressions), even when smaller wall thicknesses are used. New sealing concepts for the interface between filter element and housing, e.g. designed as a radial concept (Figure 9), render metal compounds superfluous, and result on the one hand in reduced costs and on the other hand in increased service friendliness when replacing the filter element. The two housing sections can be connected to each other by injected snap-on hooks.

Figure 9: Radial sealing principle
1 Air-filter bellows,
2 Radially acting seal,
3 Raw-air housing,
4 Clean-air hood.



Hybrid vehicles and active noise measures

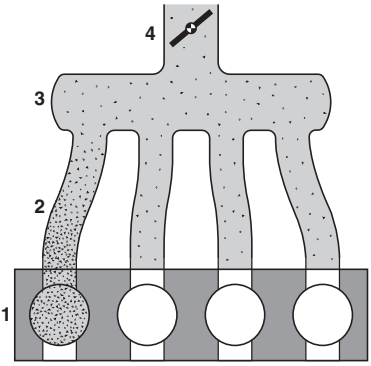
In the years to come hybridization in the automotive industry will present new challenges. Switches in operation between the internal-combustion engine and the electric motor, but also purely electric operation generate unusual acoustic states. Active systems (with ANC, Active Noise Control) constitute an interesting remedial measure in that the can not only reduce noise and thus replace passive components, but also create new degrees of freedom with regard to noise generation, e.g. for warning pedestrians in purely electric operation. Some countries have already initiated legal initiatives prescribing a warning function for purely electric operation. There are also opportunities to support a specific marque sound or for noise control in order to acoustically mask unwanted phenomena. This approach makes active systems interesting for future engine concepts.

Passenger-car intake manifolds

The main function of the intake manifold is to distribute the air homogeneously to the individual cylinders in order to ensure combustion that is as uniform and effective as possible in the individual cylinders. The quality of uniform distribution has a decisive effect on the engine's performance and emission characteristics.

In the case of gasoline engines a distinction is made between naturally aspirated engines and supercharged engines; in the case of diesel engines, on the other hand, only supercharged engines are still relevant. All three categories feature both passive and active intake manifolds. Active intake manifolds are usually called variable induction systems.

Figure 10: Ram-tube supercharging
1 Cylinder,
2 Individual ram tube,
3 Plenum chamber,
4 Throttle valve.



Intake manifolds for naturally aspirated engines

In these types of intake manifolds gas-dynamic supercharging effects are used to increase the air mass in the cylinder and thus the engine output or torque. There are two different supercharging effects here.

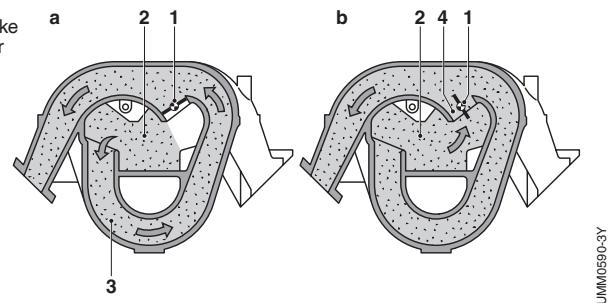
Ram-tube supercharging

Ram-tube supercharging is based on the principle that each cylinder has its own ram tube that feeds into a common plenum chamber, also called the intake plenum (Figure 10).

The intake pulse in the induction stroke triggers a vacuum wave. This propagates in the ram tube, is reflected at the opening to the intake plenum, and returns as a pressure wave to the combustion chamber. Because pressure waves always propagate at the speed of sound, the length of the ram tubes can be used to determine the engine-speed range in which the supercharging effect is wanted.

To be able to utilize this supercharging effect in numerous engine-speed ranges, there is the option of realizing active intake manifolds of this type (variable induction system, Figure 11). The optimum solution

Figure 11: Variable induction system
a) Intake-manifold geometry with changeover flap closed (torque setting),
b) Intake-manifold geometry with changeover flap open (power setting).
1 Changeover flap,
2 Plenum chamber,
3 Long, thin oscillatory intake passage with changeover flap closed,
4 Short, wide oscillatory intake passage with changeover flap open.



would be an infinitely variable induction system in which an optimum ram-tube length is assigned to each engine-speed range. But this is barely feasible, both technically and economically. For this reason mostly two- or in rare cases also three-stage induction systems are on the market.

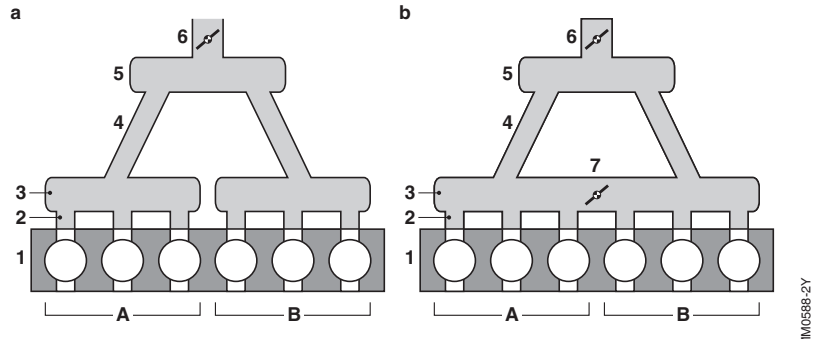
Tuned-intake-tube charging
Tuned-intake-tube charging is based on the principles of a Helmholtz resonator. Two groups of cylinders with the same ignition interval are established. The reason for this is that the intake pulses within the groups must not overlap. These groups are connected by short intake manifolds to the two resonance chambers (resonance plenums) (Figure 12). The two resonance plenums are connected by resonance (tuned) tubes to the common main intake plenum. The periodic intake pulses induce the gas oscillations to resonate at a specific engine speed and thereby generate the supercharging effect. The engine-speed range in which the supercharging effect acts is determined by the length and the diameter of the tuned intake tubes. This principle is preferably used in 3-, 6- and 12-cylinder engines because it is relatively easy, because of the firing order, to create independently operating cylinder groups (each consisting of three cylinders with 240° ignition interval).

Incorporating a resonance flap between the two resonance chambers creates an active system that provides further advantages. The intake manifold with closed resonance flap generates a torque increase in the lower engine-speed range. In the upper speed range this valve is opened to create an intake manifold with short ram tubes which increases power output in this speed range.

Intake manifolds
for supercharged gasoline engines
Intake manifolds for supercharged gasoline engines are in most cases passive intake manifolds. Power and torque are primarily determined via the charge-air pressure generated by the charger and thus by the air mass introduced into the cylinders. The still available, but mostly very short ram tubes are typically used for passive intake manifolds.

Active systems for this engine type have turbulence flaps (also called tumble flaps). These are situated near the cylinder head (Figure 13) and have the function of improving combustion in part-load operation. This arrangement firstly increases torque and secondly reduces emissions. In part load the flap is closed

Figure 12: Tuned-intake-tube charging
a) Resonance induction system, b) Resonance variable induction system.
1 Cylinder, 2 Short intake manifold, 3 Resonance chamber,
4 Resonance tube, 5 Plenum chamber (main intake plenum),
6 Throttle valve, 7 Resonance flap.
A, B Cylinder groups.



and generates a cylinder flow about the horizontal axis (tumble). In the power setting the flap is opened and the engine receives through the opening cross-section as much air mass as possible into the cylinder.

A special form here is the trough flap, which in the open setting disappears completely into the duct wall and thus ensures unrestricted flow into the cylinder.

Intake manifolds for supercharged diesel engines

In the case of intake manifolds for supercharged diesel engines there is no such clear distinction between passive and active variants. These intake manifolds practically do not have manifolds any more. All that remain are short inlet tulips that connect the intake plenum to the cylinder head.

In the case of active intake manifolds the cylinder head always has two intake ducts per cylinder, where one is a swirl duct and the other a filling duct (Figure 14). The filling duct is fitted in the intake manifold with a duct shutoff flap. The function of this flap is to fill the cylinder via the swirl

Figure 13: Tumble arrangement
a) Intake manifold in tumble setting,
b) Intake manifold in power setting.
1 Intake manifold,
2 Turbulence flap (tumble flap).

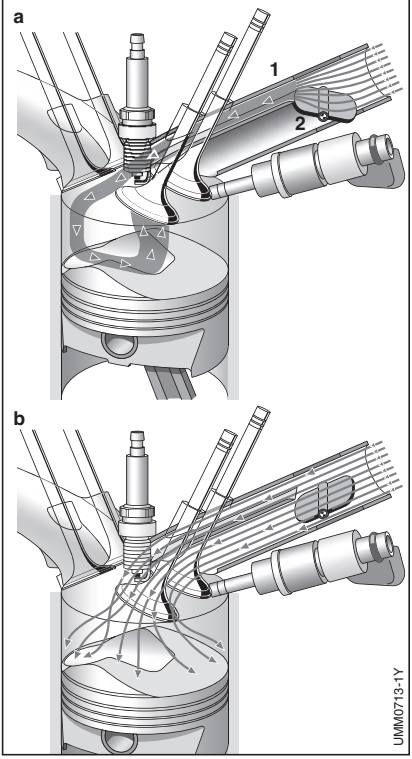
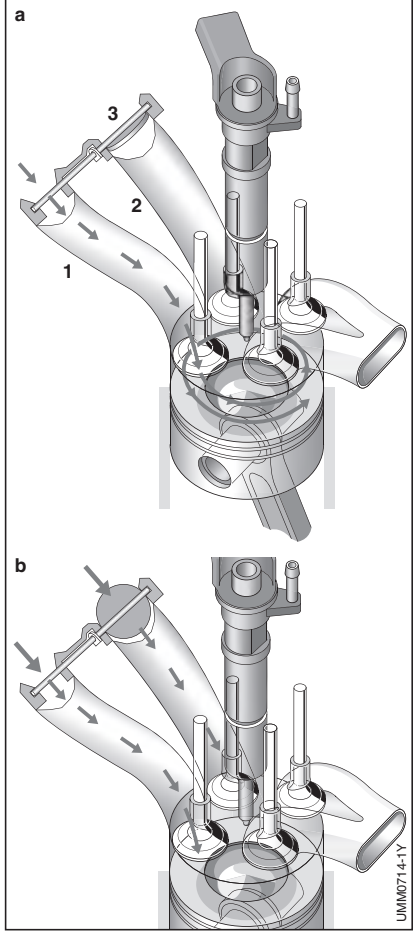


Figure 14: Swirl arrangement
a) Intake manifold in swirl setting,
b) Intake manifold in power setting.
1 Swirl duct,
2 Filling duct,
3 Duct shutoff flap.



duct in part-load operation and thus create a pronounced swirl about the vertical axis. This swirl ensures that air and fuel are thoroughly mixed and thereby helps to reduce emissions. In the transition to full-load operation the flap is then opened incrementally and the swirl is nullified by the filling duct. This results in as much air mass as possible being delivered to the cylinder and thus in high torque and high power output.

Actuator and sensor technology
Primarily electrical actuators or vacuum units are used to drive the different switching elements.

Electric actuators
Electric actuators are preferably used to activate swirl flaps in diesel engines. They enable any intermediate settings to be accurately and rapidly approached. Control is provided in most cases by the engine ECU.

Vacuum units (pneumatic actuators)
A vacuum unit can be used for pure two-point switchings, as are sufficient for length, resonance, and tumble switchings. These are switched by means of an electric changeover valve. The prerequisite is that the engine or a vacuum pump makes sufficient vacuum available.

Sensor technology
The sensor technology that features in variable induction systems has increased significantly in response to the OBD II legislation, which states that all exhaust-gas-relevant switching organs must be monitored by the engine ECU and that the driver must be alerted if they malfunction. Exhaust-gas-relevant switching elements are the turbulence flaps (tumble flaps in gasoline engines, duct shutoff flaps in diesel engines). Switching elements for length or resonance switching are not included.

For electric actuators the internal angle-of-rotation sensors of the actuators can be used in most cases. For vacuum units a separate sensor is required, but this can also be integrated in the vacuum unit.

For active intake manifolds, in addition to the sensor technology for angle-of-rotation monitoring, pressure, temperature, or combined pressure/temperature sensors are mounted on the intake manifolds. These are necessary to supply the engine ECU with the information from the intake manifold required for combustion.

Introduction of gases
A further function of the intake manifold is to distribute introduced gases uniformly into the respective cylinders. In gasoline engines these are primarily crankcase-ventilation and tank-ventilation gases. Fittings are attached to the intake manifold for these gases and then connected by the engine manufacturer to the relevant hoses. When crankcase blow-by gases are introduced, an additional heating tube may be required to avoid icing.

In diesel engines it is important for high-pressure exhaust-gas recirculation to be correctly introduced. Often uncooled exhaust gas is introduced into the intake manifold during the cold-start phase or diesel particulate-filter regeneration. Suitable shielding measures must be taken here where plastic intake manifolds are used. A central introduction point with adequate mixing section is beneficial to the extremely important uniform distribution here. The uniform-distribution requirement is very difficult to achieve in cylinder-selective introduction systems where the gas to be introduced (e.g. from exhaust-gas recirculation) is introduced via a distributor line with a separate introduction point for each cylinder (i.e. each cylinder is supplied with gas via a separate introduction point) at the very different operating points in the engine map.

Gaskets
A trend towards higher-quality materials can be discerned in gaskets. This is mainly because low-pressure exhaust-gas recirculation is more frequently used in diesel engines and this results in an aggressive condensate in the induction system. Peroxide-cured fluorocarbon rubber, for example, must be used in such applications. Due to the fact that the operating temperatures can extend both up and down, sometimes special low-temperature materials are required.

Validation of intake manifolds
Virtual and experimental product validation plays an increasingly important role for the intake manifold. The foundations for a flow- and weight-optimized component are already laid in the early development phase by the close interplay between design, testing and simulation. The optimum uniform distribution of air to the individual cylinders is optimized by CFD simulations in a number of iterative loops. The intake-manifold geometry is optimized with the aid of FEM simulations in such a way that maximum rigidity and strength are achieved with minimum weight. This enables the use of materials for the component to be kept as minimal as possible. The manufacturability of the component is analyzed and optimized with the aid of injection-molding simulations.
These theoretical analyses are then verified on sample and standard parts. The uniform distribution of intake air to the individual cylinders is measured on a flow-rate test rig. The structural strength of the components is ensured by rigorous pressure-pulsation and vibration tests under temperature influence. In the case of variable induction systems the durability of the switch-over elements and actuators is substantiated with endurance and fatigue tests.

Development trends for intake manifolds
Due to the switch from direct to indirect intercooling and the resulting benefits with regard to CO_2 emissions, the main trend in the intake-manifold market is towards the integration of intercoolers.

This trend has above all an effect on the intake manifolds in supercharged gasoline engines. Virtually all manufacturers pursue solutions involving full integration of the intercooler in the induction system. The cuboid-shaped intercooler must be integrated in the plastic intake manifold and, despite the large flat surfaces and sometimes for plastic extremely high temperatures (upstream of the intercooler), satisfy the requirements with regard to pressure-pulsation resistance. Materials that meet these requirements are currently in development.

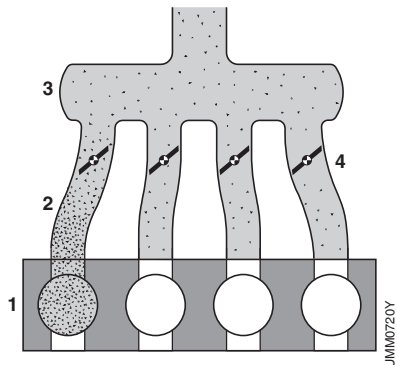
The diesel-engine market is split two ways: firstly the intercooler directly integrated in the intake manifold, and secondly the arrangement of the intercooler as a “single box” upstream of the throttle valve.

Motorcycle intake manifolds

Essentially, intake manifolds for motorcycle engines have the same functions as those for passenger-car engines. However, they do have the following particularities in terms of design: While in passenger-car applications one joint throttle valve is for the most part used for a number of cylinders (see Figure 10), for motorcycles frequently individual throttle valves are used (Figure 15). This permits a significant reduction of the intake-manifold volume between throttle valve and intake valves. The smaller volume increases the dynamics and thus the response in the event of both sudden throttle-valve opening and throttle-valve closing. The dynamic response is an elementary requirement in motorcycles. The geometry permits a smaller ratio of throttled intake-manifold volume to individual-cylinder displacement. This results in, aside from an improved response, intake-manifold pressure dynamics (see Cylinder charge).

The individual throttle valves can in this design be mounted on a single shaft or on a number of shafts (e.g., in V-engines). Comparable intake-manifold configurations are used in passenger-car engines, e.g., in racing engines.

Figure 15: Motorcycle intake manifold.
1 Cylinder, 2 Individual intake tract,
3 Plenum chamber, 4 Throttle valve.



Commercial-vehicle air-intake system

Air duct

The design of the air-intake system for commercial vehicles is essentially comparable to that of passenger-car air-intake systems (see Figure 1). What must be taken into account here is the much higher service life of commercial vehicles, the higher rated volumetric flow, and the broad range of uses in local, long-distance and construction-site applications. The different volumes and types of dust must be taken into account in filter design.

Raw-air duct, air filters and clean-air pipes up to the turbocharger are made of thermoplastics. After the turbocharger for the most part come (still) combinations of aluminum and rubber parts. However, more plastic parts will be used in future for charge-air pipes and intake manifolds (boost manifolds).

Raw-air intake

The positioning of the intake point is extremely important. A roof or side intake is usually chosen on heavy commercial vehicles. Both are generally mounted with rubber mounts for acoustic isolation to the cab rear wall. The dust and water concentrations in the air are lowest with the roof intake. There are however many raw-air duct variants to cater for the different driver's-cab heights and widths. A good compromise in terms of costs is the side intake (Figure 16): Here air is still inducted from height at the side rear of the driver's cab, irrespective of the different roof heights.

A low-height front intake or side intake, along similar lines to that used in passenger cars, is usually chosen for light utility vehicles. To keep the volumes of inducted dust a water in check, air is inducted through baffles on the vehicle sides. An air duct that delivers the best flow possible must be provided in all cases. Hot engine waste air must not be inducted as this mean a performance drop and higher fuel consumption.

Because most truck cabs are tilted forwards to access the engine for servicing, the raw-air duct is connected via bellows

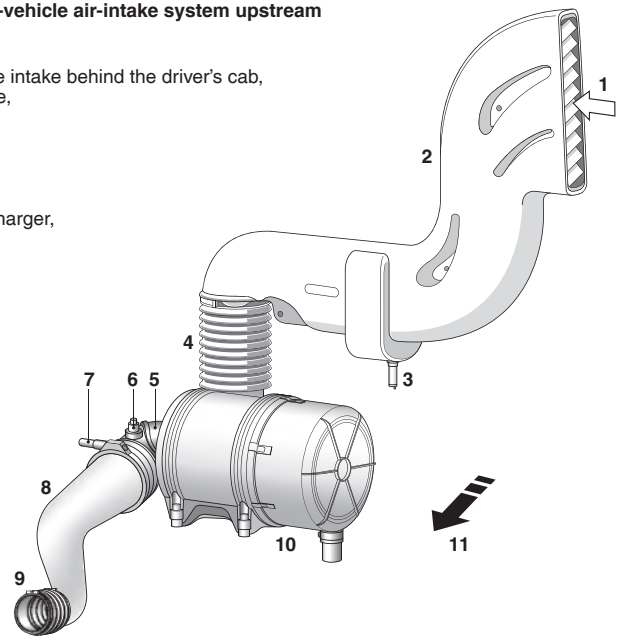
to the rest of the system. The function of the bellows is to seal the interface when the cab is tilted back and to compensate the vibrating motion of the sprung cab.

Water separation
Measures for water separation are integrated in the raw-air system. Water separation prevents dust from being washed through the filter medium, which can increase engine wear, but also cause sensors to malfunction. Because as a rule cellulose-based filter media are used, at worst the filter element may be destroyed. Measures for water separation include for example using a large intake cross-section to reduce the intake velocity so that basically as little water as possible is inducted. Baffles in the system cause the water droplets in the intake air to be thrown against the inner manifold walls, where they form a film. The water droplets are then separated via a peeling collar or a change in volume and discharged via a water-discharge valve.

Dust preseparation
In vehicle applications with a higher dust concentration in the air (e.g. construction sites or operation in dusty countries) a dust preseparator is installed in the raw-air supply, either as an individual cyclone (Figure 17) or as a cell cyclone. Cell cyclones are small cyclones which are integrated in a housing to save space. These separate out a high proportion of the coarse dust contained in the intake air. The air is induced to rotate at the inlet by guide vanes; the heavy dust particles are thrown against the inner wall of the housing and discharged through the dust outlet. The precleaned air is directed by a further guide-vane assembly (optional), which rectifies the air flow again, to the air filter.

Air filters
The most widely used air filters in commercial vehicles at present are single-stage round air filters with spiral vee-shaped filter elements. They are characterized by

Figure 16: Commercial-vehicle air-intake system upstream of the turbocharger
1 Inlet grill,
2 Raw-air duct with side intake behind the driver's cab,
3 Water-discharge valve,
4 Bellows,
5 Rubber elbow,
6 Maintenance switch,
7 Screw clamp,
8 Clean-air pipe,
9 Connection to turbocharger,
10 Air filter,
11 Direction of travel.



low pressure loss, high dust capacity, and thus long service intervals. Depending on how they are used, they can extend a typical oil-change interval by a factor of two. A further advantage of round filters is that they can also be fashioned as two-stage filters. In these variations dust separation takes place inside the air filter itself (Figure 18). The tangential air inlet generates a swirl in the housing which separates out the coarse particles.

Depending on the application, different filter media are available today which are optimized for example for maximum filtration efficiency of fine dust and soot.

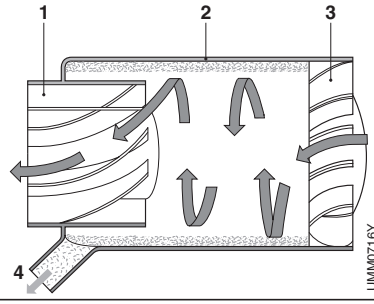
A safety element is optionally fitted inside the main filter element which prevents dust from reaching the clean-air side during servicing. It also protects the engine even if the main filter element has been damaged during servicing. Basically, however, air-filter elements should be replaced and not cleaned.

To optimize vehicle acoustics, the air filter is also fitted where necessary with resonators to minimize the opening or engine-braking noise.

Clean-air line as connection to the turbocharger

The clean-air line is as a rule a combination of metal, plastic, and rubber parts. Sensors for air mass, vacuum and air moisture can be optionally integrated. It is important for all the interfaces on the clean-air side to be permanently tight. The vacuum switch indicates to the driver that the servicing limit for the filter element has been reached. Flexible rubber connections compensate the relative movement between the engine and the frame-mounted air filter.

Figure 17: Dust preseparator (cyclone)
1 Outlet guide vanes, 2 Housing,
3 Inlet guide vanes, 4 Dust outlet.



Air filtration

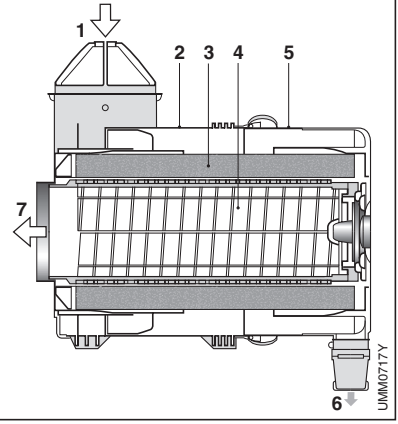
Filter elements which incorporate the latest technology achieve total mass filtration rates of up to 99.95 %. Because the service life of a commercial-vehicle engine far outstrips that of a passenger-car engine, the filtration-efficiency requirements are greater accordingly.

The filter elements have to be replaced at the intervals specified by the vehicle manufacturer. Usual replacement intervals are, depending on the application, 40,000 to 320,000 km or two years. In most cases a service-interval indicator monitors when the service interval is reached.

The media used in commercial-vehicle air filters are comparable to filter media in passenger cars, but are characterized by higher filtration efficiency and are usually designed as surface-type filters. The separated particles thus precipitate mainly on the surface of the cellulose-based medium.

Better values than with pure cellulose-based media are obtained with special nanofiber-based filter media, in which

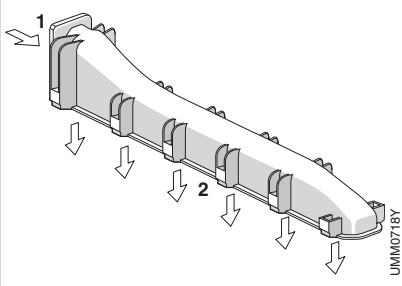
Figure 18: Commercial-vehicle air filter
1 Raw-air inlet, 2 Housing,
3 Filter element, 4 Safety element,
5 Cover, 6 Dust-discharge valve,
7 Clean-air outlet.



ultrathin, synthetic fibers with diameters of just 30 to 40 nm are applied to a relatively coarse backing layer of cellulose. The filtration-efficiency figures improved in this way enable such systems to satisfy the constantly increasing requirements of the commercial-vehicle manufacturers regarding the protection of sensors, turbochargers and other components, and by protecting these components help to comply with the ever more stringent exhaust-emissions standards.

Further development
The space allotted to the air-intake system is becoming increasingly limited, particularly due to the comprehensive exhaust-gas treatment system. For this reason, there will in future also be air filters that are flatter than round filters. This opens up the possibility of new installation positions in the vehicle, for example in front of the front wheel or underneath the driver's cab.

Figure 19: Boost manifold for commercial vehicles
1 Flange for EGR valve or auxiliary heater,
2 Flange for cylinder head.



Commercial-vehicle intake manifolds

Supercharged diesel engines require one intake manifold, usually referred to simply as a boost manifold. The air passes through the air filter to the turbocharger, and from there through the intercooler to the boost manifold. This distributes the air to the cylinders. For the most part the valve for EGR introduction (exhaust-gas recirculation) and possibly an electric auxiliary heater are flange-mounted at the inlet. The auxiliary heater helps to reach the required exhaust-gas values as quickly as possible during engine starting.

The introduced exhaust gas is mixed with the fresh air in the boost manifold and distributed as uniformly as possible to all the cylinders. The exhaust gas cooled by the EGR cooler contains, depending on the operating state and EGR rate, sulfurous acid and soot. The aluminum used for the boost manifold must be protected against the acid by for example painting. If the boost manifold is made of plastic, the plastic must be chosen so that temperature, charge-air pressure and acid do not damage the components over their service lives. The same requirements also apply to the flange gaskets.

There are both simple boost manifolds (Figure 19) which can be manufactured as a single plastic injection-molded part and also parts which due to their more complex geometry are manufactured in two parts and welded. Charge-air pressure, size of the part and tolerances must be taken into account when configuring the welding. The use of plastic also contributes to the lightweight construction of the engine.

Turbochargers and superchargers

Turbochargers and superchargers for internal-combustion engines increase the air-mass flow for a given engine displacement and a given engine speed, thereby increasing the power density. These "compressors" are generally systematically divided into three different types: the mechanically driven supercharger, the exhaust-gas turbocharger, and the pressure-wave supercharger.

In the case of the exhaust-gas turbocharger (Figure 1), the drive power is generated after banking up the exhaust gas through expansion of the exhaust gas, by which part of the exhaust-gas energy is used. The exhaust-gas turbocharger has gained acceptance for use in diesel engines in both the automobile and commercial-vehicle sectors. The exhaust-gas turbocharger, particularly in conjunction with gasoline direction injection, is being increasingly used in gasoline engines as well.

But mechanically driven superchargers directly coupled to the crankshaft are also used in some vehicle models with gasoline engines. Pressure-wave superchargers on the other hand, despite having some benefits, did not enjoy widespread use in the automotive sector; only a few vehicle models were equipped with these superchargers in the 1980s and 1990s.

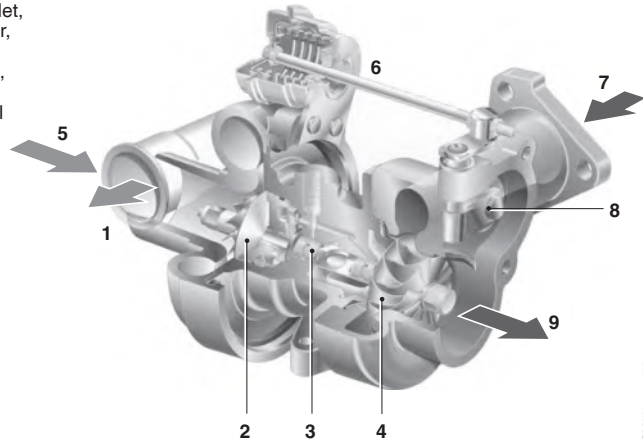
Superchargers (mechanically driven)

In the case of the mechanically driven supercharger, the required drive power is branched directly from the crankshaft via a belt or a gear drive from the engine output, i.e., the supercharger and the engine are mechanically coupled. Mechanically driven superchargers [1] fall into two categories: Those which operate according to the positive-displacement principle and those which compress air according to the flow principle and the principle of momentum. The internal-combustion engine drives the supercharger as a rule by means of a belt drive. It is driven either directly (continuous-running operation) or via a clutch.

Mechanical centrifugal supercharger
The compressor of the mechanical centrifugal supercharger operates according to the flow principle. This type of device is very efficient, providing the best ratio between unit dimensions and volumetric flow when compared with other mechanically driven superchargers. The pressure ratio that can be achieved is dependent on the peripheral velocity of the compressor impeller. The air flow rates typical of passenger-car engines give rise to small compressor-impeller di-

Figure 1: Exhaust-gas turbocharger with wastegate (view)

1 Compressed-air outlet,
2 Compressor impeller,
3 Shaft,
4 Exhaust-gas turbine,
5 Fresh-air inlet,
6 Actuator with control rack for actuating the wastegate,
7 Exhaust-gas mass-flow intake,
8 Wastegate (boost-pressure control valve, in turbine inlet),
9 Outlet, exhaust-gas mass flow.



mensions. The pressure ratios required for effective engine supercharging thus require high compressor speeds.

The range of application of the mechanical centrifugal supercharger is limited technically by the high rotational speeds required and the transmittable drive power, and economically by the comparatively high costs. It is used in small numbers in performance-oriented engine concepts.

Positive-displacement supercharger

A comparatively large number of designs has been proposed for units operating according to the positive-displacement principle, but only a few have succeeded in being employed in current series applications. Positive-displacement superchargers can operate with or without internal compression. Internal-compression superchargers include the reciprocating-piston, screw-type and rotary-piston compressors. The Roots supercharger is an example of a unit without internal compression. All of these positive-displacement superchargers share certain characteristics as shown here in the program map of a Roots supercharger (Figure 2):

– The speed curves n_L = const in the p_2/p_1-$\dot{V}$ program map are very steep, i.e., the volumetric flow $\dot{V}$ decreases only slightly as the pressure ratio p_2/p_1 increases. The drop in volumetric flow is basically determined by the efficiency of the gap seal (backflow losses). It is a function of the pressure ratio p_2/p_1 and of time, and is not influenced by rotational speed.
– The pressure ratio p_2/p_1 does not depend on the rotational speed. In other words, high pressure ratios can also be generated at low volumetric flow rates.
– The volumetric flow $\dot{V}$ is independent of the pressure ratio and, roughly formulated, directly proportional to rotational speed.
– The unit retains stability throughout its operating range. The positive-displacement compressor operates at all points of the p_2/p_1-$\dot{V}$ program map as determined by supercharger dimensions.

Mechanical positive-displacement superchargers generally must be substantially larger than mechanical centrifugal superchargers in order to produce a given volumetric flow.

Roots supercharger

The Roots supercharger is a rotary-piston machine with contra-rotating twin- or multi-bladed rotary pistons mounted in rolling bearings and synchronized by a gear set. The rotary pistons mesh at identical speed without directly contacting each other in the housing (Figure 3). The gaps between these components essentially determine the machine's efficiency.

Roots superchargers operate without internal compression. Mufflers are generally fitted on the suction and pressure sides to limit their acoustic impact. Their system

Figure 2: Program map of a Roots supercharger

p_1 Total pressure before compressor,
p_2 Total pressure after compressor,
n_L Supercharger speed.

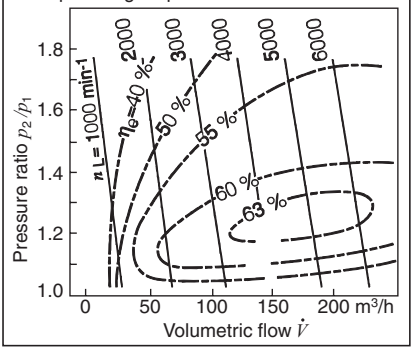
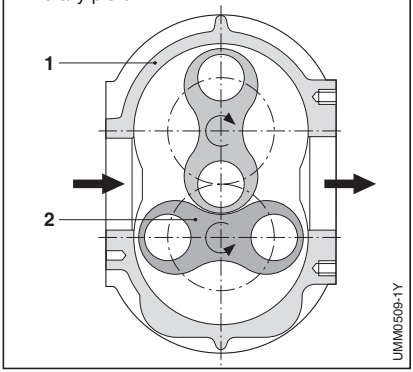


Figure 3: Cross-section through a Roots supercharger

1 Housing,
2 Rotary piston.

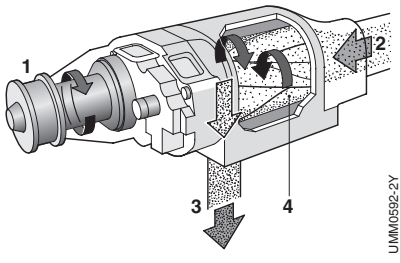


and design limits the pressure ratios that can be achieved to levels below 2. Coating the functional components has helped to improve efficiency. Current development work is focused on among other things variable-speed gearboxes on the drive side.

Screw-type supercharger
The screw-type supercharger (Figure 4) is very similar in design to the Roots supercharger, i.e., it is a twin-shaft contra-rotating rotary-piston machine. However, it differs from the Roots supercharger in that it generally operates with internal compression. It can achieve higher pressure ratios than a Roots supercharger. On the suction side (inlet) the rotation of the pistons (screws) opens a profile-gap space which fills with inducted air. As the screws continue to contra-rotate, the profile-gap space continuously decreases in size until it reaches the outlet control edges. At this point internal compression is completed and the compressed volume is forced into the pressure connection (outlet). In order to minimize internal leakage losses, the tolerances between the screws and the walls must be very closely adhered to. Measures along similar lines to the Roots supercharger were implemented for improving the function and efficiency of this type of supercharger.

Current developments are focused on the integration of electric drives primarily for fuel-cell applications. In addition to electric-only assemblies there are also variants with a power split between mechanical and electric drive.

Figure 4: Screw-type supercharger
1 Drive, 2 Air inlet,
3 Air outlet (compressed air),
4 Screws.

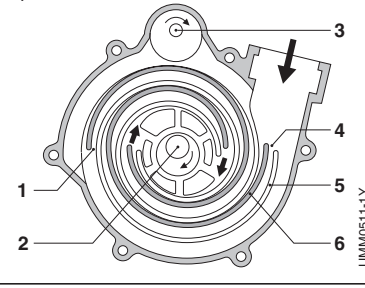


Spiral-type supercharger
The spiral-type supercharger (Figure 5) is a compressor in which a rotor with spiral lands eccentrically describes a circular path in a housing also with spiral lands. The spiral-type supercharger operates according to the principle of a circulatory displacer element. Working chambers in sequence open for charging, close for transport, and open once again for discharge at the hub.

The G-supercharger (G-Lader) is an example of this type. Maintaining very tight tolerances and the seal has a function-determining effect. Radial sealing is provided by the narrowest gaps possible, axial sealing by touching sealing strips inserted on the end faces. These sealing strips are subject to wear and are if necessary replaced during regular vehicle servicing. Internal compression can be achieved by an appropriate configuration of spirals.

More recent developments are heading towards a simplification of the design and the integration of a switchable clutch.

Figure 5: Cross-section through a spiral-type supercharger
1 Air inlet into second working chamber,
2 Drive shaft,
3 Displacer element,
4 Air inlet into primary working chamber,
5 Housing,
6 Displacer element.



Exhaust-gas turbochargers

Of all the supercharger units in use, the exhaust-gas turbocharger is by far the most widespread. While nowadays virtually 100 % of passenger-car and commercial-vehicle diesel engines are turbocharged, the proportion of supercharged gasoline engines will continue to increase dramatically in the next few years. Here, too, the exhaust-gas turbocharger will be the dominant supercharger.

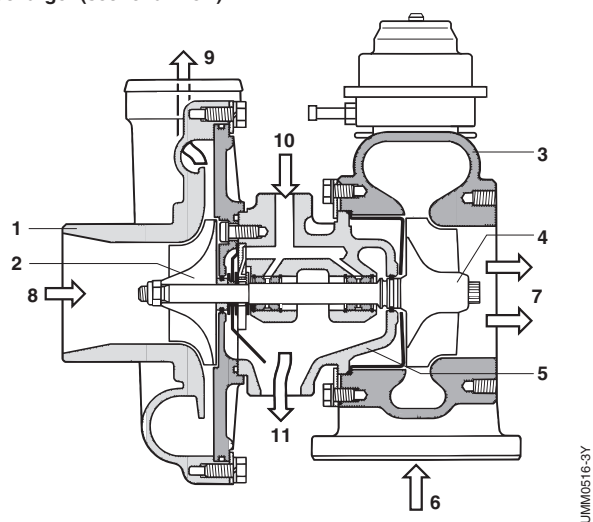
Design and operating principle
The exhaust-gas turbocharger consists of two turbo elements: a turbine and a compressor, where the impellers for the two elements are mounted on a common shaft (Figures 6 and 1). The operating concept of a turbine engine is based physically on the principle of momentum. The turbine converts some of the enthalpy contained in the exhaust gas into mechanical energy to drive the compressor. The compressor draws in fresh air through the air filter and compresses it.

The exhaust-gas turbocharger is coupled only thermodynamically, not mechanically, to the engine. Turbocharger speed does not depend on engine speed, but is rather a function of the balance of drive energy between the turbine and the compressor. The turbine generates the compressor output and the output which is essentially used up in the bearings and ultimately dissipated as heat (mechanical losses).

Applications
The exhaust-gas turbocharger in its present form can trace its roots back to the work of Alfred Büchi (1905) [2], who had already recognized the potential of combining supercharging and valve overlap to scavenge residual exhaust gas (1915). Exhaust-gas turbochargers have traditionally been used for the purpose of supercharging large diesel engines, originally being used primarily for commercial-vehicle, marine and locomotive power plants as well as for agricultural- and construction-machinery applications.

Use in diesel vehicles
The mid-1970s saw the advent of the first production diesel engines supercharged by exhaust-gas turbochargers in passenger cars. The introduction of a “wastegate” for controlling boost pressure finally facilitated a torque-oriented design and thereby improved driveability. Further increases in the performance of passenger-car diesel engines have been achieved through the

Figure 6: Exhaust-gas turbocharger (sectional view)
1 Compressor housing,
2 Compressor impeller,
3 Turbine housing,
4 Turbine wheel,
5 Bearing housing,
6 Incoming exhaust gas,
7 Outgoing exhaust gas,
8 Atmospheric fresh air,
9 Precompressed fresh air,
10 Oil supply,
11 Oil return.



introduction of direct injection (1987) and by exhaust-gas turbocharging with variable turbine geometry (1996) or two-stage exhaust-gas turbocharging (2004). These developments have been reflected in a marked increased in the market share for diesel engines in Europe. Today in Europe all passenger-car and commercial-vehicle diesel engines are equipped with exhaust-gas turbocharging and intercooling (charge-air cooling).

Use in passenger cars with gasoline engines

Exhaust-gas turbocharging of gasoline engines was originally the reserve of only high-output sports engines for increased performance and, because of inadequate driveability ("turbo lag"), was rarely encountered on the market. In the meantime, gasoline-engine supercharging has become an integral part of engine development, predominantly in connection with small to medium-sized engines. In addition to the improvement in efficiency, one of the main objectives is to avoid the increases in the number of engine cylinders (e.g., from four to six cylinders), thus positively influencing package space and fuel consumption.

In contrast to diesel engines, mechanical supercharging was also employed and today continues to be employed occasionally on gasoline engines on account of the excellent transient build-up of boost pressure. With the introduction of direct injection in the gasoline engine the exhaust-gas turbocharger was almost on a par with mechanically driven superchargers thanks to the use of scavenging.

In addition, a combination of mechanical supercharging and exhaust-gas turbocharging (combination supercharging) was in the past employed to provide small-capacity gasoline engines with high rated output and high torque at a comparatively low specific speed. However, the current trend is towards the use of electrically driven supplementary compressors instead of combination supercharging.

The exhaust-gas turbocharger with variable turbine geometry is today the standard supercharger used in diesel engines. In the past this technology was only used on gasoline engines in niche applications due to the unmastered, thermal and technical requirements and the simultaneously high component costs. Current high-volume applications, however, are showing a trend reversal with regard to this issue, which was instigated by the series launch of the 96-kW engine version of the Volkswagen EA211 evo. The technical suitability of the turbocharger with variable turbine geometry is matched to a low-emission Miller combustion process without valve overlap (scavenging) with a simultaneously increased compression ratio. In the first generation the exhaust-gas temperature is limited to 880 °C. However, current developments cover exhaust-gas temperatures up to 1050 °C.

With regard to the statutory exhaust-emission and consumption guidelines and the existing performance figures required by customers, the importance of exhaust-gas turbocharging to the reduced-capacity and reduced-cylinder engines currently in development ("downsizing") cannot be underestimated. At present the number of turbocharged gasoline engines is increasing. The growth of this market segment is expected to continue over the coming years.

Exhaust-gas turbocharger design

The exhaust-gas turbocharger essentially consists of three assemblies: the bearing housing, the compressor, and the turbine (Figures 6 and 1). Depending on the design, there may be a fourth assembly: a boost-pressure control facility.

Bearing housing

The bearing housing accommodates the bearings and the elements for sealing the shaft. State-of-the-art designs usually feature a specially developed friction bearing both in the radial as well as the axial (thrust) bearing assembly. The radial bearings are designed either as rotating double plain bushings or as stationary plain-bearing bushings. The thrust bearing in conventional designs is made up of a multiple-spline surface bushing that is subject to load from both sides.

Lubricating-oil supply
The lubricating oil is supplied by connecting the turbocharger to the engine's oil circuit. The oil outlet is connected directly to the oil pan in the crankcase. Today, this type of bearing assembly is used to control rotational speeds of more than 300,000 rpm reliably.

Bearing arrangement
A rolling-bearing rotor assembly (shaft with turbine wheel and compressor impeller) can deliver consumption-related efficiency advantages which provide for a better transient build-up of boost pressure and higher boost pressures in the part-load range. However, this trend is currently falling off because significant improvements could not be achieved with the evolutionary development of the conventional, hydrodynamic-bearing version. Modern developments have a reduced friction loss with simultaneously very good acoustic properties, as demonstrated by for example the "3D-Bearing" from BMTS-Technology GmbH and the "Z-Bearing" from Garrett Motion AG. However, rolling-bearing rotor assemblies also feature in series production in the passenger-car sector. For deeper market penetration, however, further developments with regard to costs, acoustics and service life are necessary. In the commercial-vehicle sector wider deployment is expected in long-distance haulage now that the tendency to wear of the rolling elements has been sufficiently reduced.

Sealing
To seal off the oil chamber to the exterior and to minimize the entry of charge air ("blow-by") and exhaust gas to the charger interior, the shaft is equipped at the casing openings with piston rings which are tensioned in the bearing housing and with the shaft groove form a simple labyrinth. The state-of-the-art sealing technology used today limits the permissible installation location for the exhaust-gas turbocharger to a comparatively small range of inclination. Touching seals would be useful in this respect, but have not been able to be produced up to now at reasonable cost owing to the high relative speeds between shaft and housing.

Cooling
No additional cooling measures are necessary to maintain efficient operation of the bearing assemblies under suitable ambient conditions in the engine compartment, i.e., at exhaust-gas temperatures in conventional passenger-car usage up to 830 °C. The relevant temperatures can generally be maintained below critical levels using devices such as a heat shield, and by thermally isolating the hot turbine housing, supplemented by incorporating suitable design elements in the bearing housing itself. Water-cooled bearing housings are employed for higher temperatures, for example, on gasoline engines operating at up to 1080 °C exhaust-gas temperature or in certain diesel-engine applications.

Compressor
The compressor assembly in passenger car and commercial-vehicle applications consists of an axial-inflow and radial-outflow impeller (compressor impeller, centrifugal design) and a generally guide-ring-free compressor housing made of cast aluminum. However, the use of guide rings is only practical if a limited mass-flow spread of the engine operation curve can be applied without driveability limitations. The are primarily suitable for applications with high compressor pressure conditions. The flow is further slowed down in the housing, in the course of which some of the pressure build-up occurs. Having the smallest possible contour gap between the impeller and the housing is crucially important to efficiency.

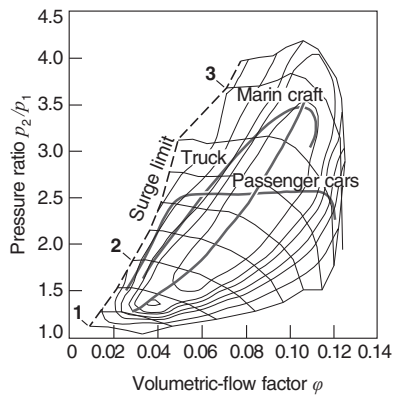
Compressor impeller

The compressor impeller can be mass-produced in a special casting process from an aluminum alloy. Today, however, impellers milled from an aluminum forging alloy are mainly used. In the event of more rigorous service-life requirements, particularly in the commercial-vehicle sector with critical application profiles, impellers milled from a titanium alloy are also used. Aside from cost-intensive titanium alloys, aluminum-lithium alloys are among the preferred materials. For engines with low-pressure-side exhaust-gas recirculation, where the compressor has to deliver medium containing exhaust gas, the compressor impeller is provided with a coating. This protects the compressor impeller against erosion by particulate and droplet impingement.

Although the compressor impeller contributes much less than the turbine wheel to the rotor assembly's mass moment of inertia, a value which is as low as possible is also important on the cold side. Because of the heightened requirements with regard to performance and service life, attempts to mass-produce plastic compressor impellers for turbocharger use have hitherto failed.

Figure 7: Compressor program map with typical engine operation curves valid for all displacements

1 Peripheral velocity u = 150 m/s,
2 Peripheral velocity u = 300 m/s,
3 Peripheral velocity u = 450 m/s.
u Peripheral velocity,
p_1 Total pressure at compressor inlet,
p_2 Total pressure at compressor outlet.
The volumetric-flow factor
is a dimensionless volumetric flow.



Compressor program map

The characteristic of a compressor is described by a program map (Figure 7). In contrast to a positive-displacement supercharger, the program map of a flow compressor is limited by gas-dynamic, kinematic and centrifugal effects. By implementing appropriate measures, the effective range as well as the speed and efficiency characteristics of the compressor can be adapted to the required boost-pressure curve. The effective range of the compressor is determined on the "left" side of program map (i.e., maximum achievable pressure ratio for a given throughput) by the surge limit. On the "right" side (maximum possible throughput, limited by the attainment of the velocity of sound in the narrowest cross-section) the choke limit restricts the operating range. The surge limit is defined as the transition from the stable to the unstable operating range. The stability limit is dependent on the vibrational dynamics of the entire conduit system in which the compressor stage is operated. If the stability limit is undershot, delivery of the air-mass flow is as a rule periodically interrupted and re-established. The extent of the surge limit can therefore be improved among other things by means of a suitable intake-line design. The position of the surge limit is therefore not a compressor property, but rather a system property. The choke limit describes the throughput limit of a flow compressor which occurs when transonic flow velocities are reached at the narrowest compressor cross-section. The choke limit is characterized by steeply falling speed curves.

In view of the proportionality of the engine speed and volumetric air flow of the compressor, it is easy to deduce that compressors for turbocharging gasoline engines must have a substantially larger effective map width than, for instance, compressors which supply large diesel engines with air. The dimensionless compressor program map in Figure 7 shows the characteristic air requirements along the WOT

for passenger cars, commercial vehicles, and large heavy-duty engines.

The aerodynamic properties of the compressor stage are determined by the geometric properties of the compressor impeller, the diffuser, and the spiral housing. Furthermore, the compressor impeller must have a suitable mechanical design. The for the most part compromised geometric design is necessary to ensure safety again the failure modes that occur. Damage is caused by high-frequency oscillation excitations and tensile loads and shear stresses resulting from the underlying, application-specific speed horizon which act on the impeller geometries.

With the appropriate, technical requirement, recirculation channels integrated into the compressor housing or a special form of compressor inlet channel (preliminary volumes) are used to extend and stabilize the compressor program map.

Recirculation channel

The recirculation channel at suitable operating points directs some of the delivered mass flow back to the compressor inlet and thereby changes the flow condition at the compressor impeller. With suitable compressor impellers, the surge limit can mainly at pressure ratios greater than 2 be shifted to smaller mass flows. In addition, the bypass in the area of high mass flows results in a slight increase in throughput in the single-digit percentage range. The achievable efficiency levels decrease as a rule when a recirculation channel is used.

Inlet and discharge guide vanes

Variable inlet or discharge guide vanes are also used to extend the program map. The program map can also be extended in the surge-limit range by this. This means that the compressor can be operated at an operating point at which the surge limit would already be exceeded. Apart from a simple preliminary diffuser, whose vane angles twist into a specific position as the mass flow increases by means of the flow forces and are released again as the mass flow decreases, variable compressors have up to now not been series-produced.

Variable compressors

Still in development are variable compressors in which a variable cross-section constriction is arranged upstream after the narrowest compressor cross-section. This enables the surge limit to be shifted towards lower mass flows. This is consequently a further map-extending or an efficiency-increasing measure for operating points close to the surge limit.

Blow-off valve

To date compressor variants for turbocharging conventional gasoline engines are equipped with a blow-off valve, whose task it is to prevent compressor surge when load is quickly removed (i.e., throttle valve closes). The blow-off valve, which was originally actuated pneumatically but is now mainly actuated electrically, for this purpose creates a bypass between the compressor outlet and inlet. In this way, the compressor pumps briefly as a blower in the circuit. Today it is possible, above all by introducing electrical actuation of the wastegate, to prevent surge without the use of a blow-off valve and guarantee reliable operation.

Turbochargers with variable turbine geometry, opening up new cutoff-strategy possibilities, are increasingly gaining acceptance in current engine applications with modern Miller combustion processes.

Turbine

The turbine consists of a diffuser and a (turbine) wheel. The diffuser of the turbocharger turbine is integrated in the flow housing (spiral housing) and for passenger-car applications, if the turbocharger is one without variable turbine geometry, is designed without vanes. Fixed-geometry diffusers with vanes are frequently used in exhaust-gas turbochargers for large engines for better fine graduation of the turbine throughput. The flow is accelerated in the diffuser and distributed as evenly as possible to the turbine wheel.

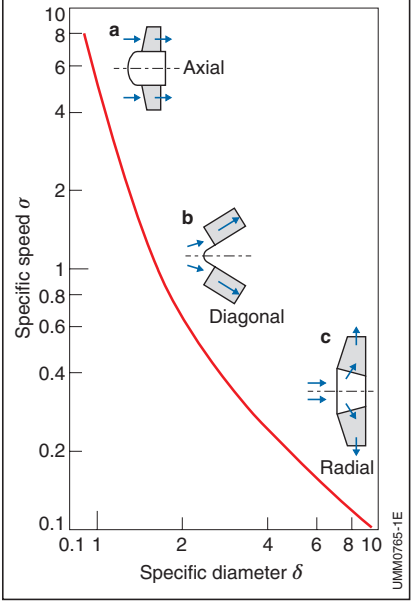
Turbine wheel

The turbine wheel is normally designed as a centripetal turbine, i.e., radial inflow and axial outflow.

Because of their favorable mass moment of inertia, turbine wheels of semi-axial design are also used. These have diagonal inflow and axial outflow. But these are not so-called diagonal turbines, which are unsuitable for these too low, specific throughputs.

Axial wheels, which have axial inflow and axial outflow, are only installed in turbochargers for medium-speed and slow-running large engines. The achievable pressure ratios are greatly limited owing to the prevailing short channel routing here with a comparatively high, specific throughput. Axial wheels have a comparatively high speed number, a dimensionless quantity that is obtained from the quotient of the peripheral velocity and the flow velocity of the turbo-machine. Without suitable matching only axial compressors can be paired with axial turbines in each case; the same applies to radial turbines.

Figure 8: Cordier diagram
a) Turbine wheel, axial design,
b) Turbine wheel, semi-axial design,
c) Centripetal turbine.



In special cases multi-flow centripetal machines are paired in order to reach a sensible level for the corresponding speed number. The interrelationship of specific mass flow and suitable design can be inferred in many available representations of the so-called Cordier diagram (Figure 8, [3], [4]). The speed number is plotted against the diameter number in the Cordier diagram. The diameter number is a dimensionless quantity for characterizing turbo-machines [5]. It relates the outside diameter of the machine impeller to the calculated diameter of an equivalent nozzle.

Exhaust-gas pipe

The arrangement of the exhaust-gas pipe plays an important role in supercharging. Classically, a distinction is made between pulse supercharging and constant-pressure supercharging. In the case of pulse supercharging, the exhaust-gas pipes are routed separately to the turbine housing. Those cylinders whose exhaust pulses influence each other minimally are brought together on the exhaust-gas side. This means that the scavenging processes of the cylinders that are brought together are subject to less hindrance. The turbine housing is designed in such a way that the separation of channels is maintained as far as possible up to the turbine-wheel inlet. The channel cross-section is the size of the exhaust cross-section. The exhaust pulses reach the turbine inlet in the form of pressure waves such that pulsating pressure is applied to the turbine. This effect must be utilized at low engine speeds where the pressure surges, due to the long interval from one exhaust pulse to the other, make up substantial power contribution of the available exhaust-gas enthalpy.

In the case of constant-pressure supercharging, the exhaust discharges are directed in a common, comparatively large-volume collector to the turbine housing. This process smooths out the individual exhaust pulses to a large extent.

With constant-pressure supercharging the more uniform enthalpy over time can be processed with greater efficiency. Pulse supercharging, on the other hand, offers advantages for the engine's part-load and accelerating performance.

Turbine housing

The design of the turbine housing differs significantly, depending on whether constant-pressure or pulse supercharging is used. Pulse supercharging for the most part is used in commercial-vehicle applications. The turbine housing is designed as double-flow housing in which the two flows join only just before the wheel inlet.

Constant-pressure supercharging is used on high-speed engines such as the diesel engines in passenger vehicles. The turbine housings are single-flow ("single-scroll") in design and partially integrated into the exhaust manifold as a single component. This provides for a particularly compact component geometry.

Gasoline engines utilize both constant-pressure and pulse supercharging. The exhaust-gas systems are configured accordingly and the housings are single-flow or twin-flow ("twin-scroll" or "dual volute") in design. Partially manifold-integrated turbine housings are also used in exhaust-gas turbochargers for gasoline engines.

Materials

The turbine wheels are made of high-alloy nickel steels on account of the high demands imposed by temperature and mechanical loads. In view of the accuracy of shape required, procedures such as vacuum or centrifugal casting are used. The mechanical connection with the shaft is achieved by means of friction, laser or electron-beam welding.

Non-metallic and ceramic materials were analyzed as an alternative for turbine wheels. The material class titanium aluminide (TiAl) appeared promising. Due to the low density the mass moment of inertia is reduced to approx. 50 % compared with common materials like IN713C (nickel-base casting alloy). Unfortunately, the available alloys exhibited to a large extent unsuitable material properties, as a result of which the associated expensive development activities were discontinued.

The associated turbine housings are made of different casting materials depending on the thermomechanical requirements. Exhaust-gas temperatures in real applications in passenger cars range between approx. 750 °C and approx. 1100 °C. Here the spectrum of casting materials used ranges from cast iron to high-alloy cast-steel variants. Aside from the marked focus on cost, the designs are characterized by the smallest possible thermal inertia. Water-cooled turbine housings made of wrought aluminum alloys which can be operated up to an exhaust-gas temperature of 1030 °C constitute a special solution. Turbine housings for dual-flow turbines consist primarily of cast-steel materials because these alone have the requisite thermomechanical properties. Turbine housings built up from welded sheet-metal parts in single-walled or twin-walled form have also been produced in AGI versions (AGI: Air-Gap-Insulated). AGI components reduce wall heat losses and have a low thermal inertia.

So far built-up turbine housings have only been introduced in series production to a limited extent. However, these could help to meet future emission regulations in that they enable the catalytic converter to be heated up more quickly after cold starting.

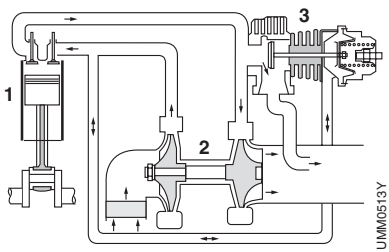
Boost-pressure control

In view of the large rotational speed spread of passenger-car engines, a boost-pressure control facility is indispensable to maintain the maximum permissible boost pressure for an acceptable design torque. Standard practice presently favors turbine output control on the exhaust-gas side.

Bypass control
A simple and in gasoline and commercial-vehicle engines widely used method is that of bypass control, in which generally some of the exhaust-gas flow is routed through a flap valve ("wastegate") around the turbine (Figures 1 and 9). The flap valve is actuated by a pneumatic unit or an electric actuator. The pneumatic unit can take the form of an overpressure unit (supplied by the boost pressure itself) or a vacuum unit (vacuum supply from the vehicle system). If the boost pressure is used, the wastegate cannot however be actuated independently of the engine operating state. The control pressures are usually adjusted by using clocked-pulse valves. Because of the thermal load, the actuators are generally mounted on the compressor side and connected via a control rack to the wastegate lever.

Current developments involve almost exclusively electric actuators with position feedback, which provide for a fast and precise control response. The wastegate can therefore be actuated independently of the engine operating state. This is necessary to comply with the current statutory requirements with regard to emissions and consumption. Currently available electric actuators impress with their compact design and high actuating forces, which guarantee a better seal seat for the wastegate valve (overrun control valve) and an accelerated build-up of boost pressure.

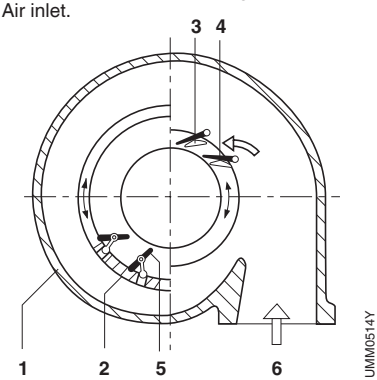
Figure 9: Boost-pressure control via exhaust-side boost-pressure control valve (wastegate)
1 Engine,
2 Exhaust-gas turbocharger,
3 Boost-pressure control valve (wastegate).



Variable turbine geometry
Compared to bypass control, variable turbine geometry (VTG) offers a more efficient method of boost-pressure control over the entire operating range. Today it is state of the art in passenger-car diesel engines and will in future also be introduced in commercial-vehicle and gasoline-engine applications. The entire mass exhaust-gas flow is routed here via the turbine, providing specific benefits in exploiting the available energy.

Of all the potential designs, adjustable blades have achieved general acceptance, as they combine a wide control range with high efficiency levels (Figure 10). The blade angle can be easily adjusted by means of a swiveling movement. The blades, in turn, are set to the required position using adjusting cams, or directly via control levers attached to the individual blades. All blades engage in an adjusting ring which can be turned by means of a lever. The lever is actuated by pneumatic or electric actuators. Where a pneumatic unit is fitted, today exclusively a position sensor is integrated in the unit, and the signal from this sensor ensures that the VTG actual blade position is clearly identified.

Figure 10: Variable turbine geometry (schematic)
1 Turbine housing,
2 Adjusting ring,
3 Control cam,
4 Adjusting blades (pivot point at front edge of blade),
5 Adjusting blades (pivot point at approx. 50 % of blade length),
6 Air inlet.



The first design variations had symmetrical blades with straight median lines. Today, profiled blades with pronounced pressure and suction sides and curved median lines are used. This helps to improve efficiency above all in the closed position, i.e., when accelerating from lower full load and in the part-load range. It is particularly important for engine performance in the lower engine speed range to adhere to the smallest possible axial gaps between the adjustable blades and the adjacent components.

Owing to the high exhaust-gas temperatures of a gasoline engine, the operationally reliable and durable implementation of an adjusting-blade exhaust-gas turbocharger involves intensive matching of thermomechanics, choice of materials, and kinematics. For this reason an adjusting-blade VTG was for a long time only available in just one gasoline-engine application of one sports-car manufacturer. In 2016, however, this technology also went into series production in mid-size passenger cars with gasoline engines. Today they also exist in price-sensitive vehicle segments in Europe and China.

A Japanese manufacturer developed, aside from adjustable-blade VTG, a comparatively simple variability in the area of the turbine spiral. However, to date it has not attained the performance levels of adjustable-blade VTG.

There is a further operating principle, the so-called sliding-sleeve principle. This is a compromise solution which generates a comparatively high degree of turbine efficiency with fewer hot parts than the adjustable-blade variant. A single actuator is used to achieve both variability of the turbine mass flow and the operation of an internal bypass to circumvent the turbine.

Exhaust-gas turbochargers with supply of additional energy

Modern exhaust-gas turbochargers are characterized by high efficiency levels and a low mass moment of inertia on the part of the rotor assembly such that a very good transient build-up of boost pressure can be achieved, particularly in conjunction with measures on the engine, such as for instance “scavenging” [6]. However, in view of current legislation, this measure is still only occasionally used because it can lead to a significant oxygen content in the exhaust gas. In combination with gasoline-engine combustion processes and the associated exhaust-gas temperatures, an oxygen excess can lead to accelerated aging of the catalytic-converter substrate and consequently to a reduction of the guaranteed conversion rate to be expected.

In the past various solutions have been proposed in order in states with little available exhaust-gas energy to supply additional mechanical, hydraulic or electrical power to the shaft for the purpose of accelerating the rotor assembly more quickly. For example, the incorporation of a gear on the turbocharger shaft allows mechanical energy to be supplied via a gearbox and a switchable mechanical connection to the crankshaft.

A further approach provides for installing on the rotor assembly between the bearings a Pelton turbine wheel which is pressurized with high-pressure oil (approx. 100 bar) from the engine circuit or separately with hydraulic fluid.

Electrically assisted exhaust-gas turbocharger

In a further approach it is proposed to install on the turbocharger shaft, for instance between the bearings, an electric motor which is supplied via power electronics by the vehicle electrical system (“electrically assisted exhaust-gas turbocharger”). The electrically assisted turbocharger is able in certain operating phases to function as a generator and thereby return electrical energy to the vehicle electrical system.

The current issues of consumption and emissions have thrown up an important additional task. The unacceptable aging of exhaust-gas treatment creates the additional requirement of high expansion performance on the part of the turbine stage at full-load speeds. The aim here is to lower the exhaust-gas temperature downstream of the turbine to as low a level as possible so that the exhaust gas permanently guarantees the conversion of the catalytic converter. A reduction of the exhaust-gas temperature to values lower than 900 °C would be desirable. This requires a particularly precise matching of the throughput and efficiency offering. A turbine bypass will not resolve the issue here, or must be able to be effectively switched off.

There are currently a number of variants of the electrically assisted turbocharger in development. There are variants without variability and without bypass valve. The recuperation potential exceeds the excess exhaust-gas enthalpy offering, by which the electrically assisted turbocharger can be braked at any time to the desired operating speed. However, this only applies to full load; a constantly sufficient electrical acceptance is the prerequisite here. This solution is a compromise which is intended to make use of a simple, cost-effective and fail-safe fixed-geometry exhaust-gas turbocharger.

Alternatively, the turbine has a wastegate so that even light-off operating modes (catalytic-converter heating-up phase until the targeted conversion of pollutants from the exhaust gas is achieved) can be supported by the turbocharger itself. Furthermore, there are electrically assisted turbochargers with variable turbine geometry. These have the distinction of being able to react much more freely to the equilibrium between available exhaust-gas enthalpy, energy recuperation demand and boost-pressure demand. With a suitably designed turbocharger a number of operating modes can be simultaneously run and can also be used in line with demand.

Compared to a conventional turbocharger, a magnet arrangement mounted on the shaft results in an increased moment of inertia on the part of the rotor assembly. The potential of these approaches does make it possible to shift the turbocharger's power balance independently of the engine, but this is limited by the program-map limits of the compressor used. Suitable measures for shifting and extending the compressor-map limits are set out in the section entitled “Compressor”.

The first series launches with the use of an electrically assisted turbocharger in passenger cars is expected from 2021.

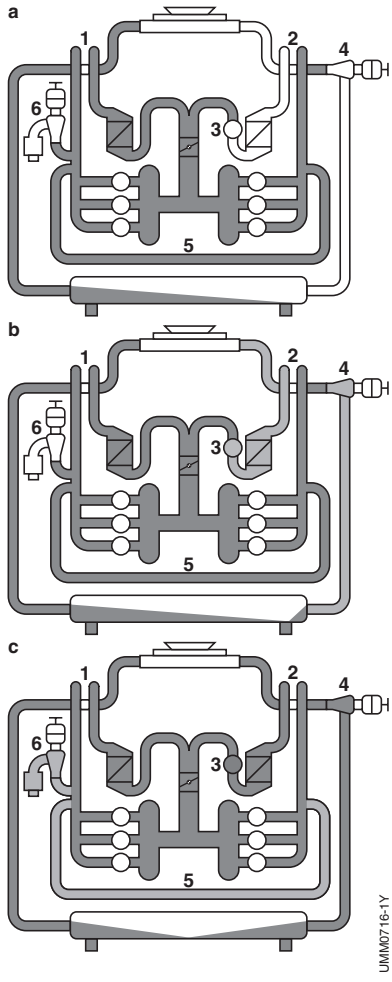
Complex supercharging systems

With a parallel or serial/multistage connection of supercharger units the power limits can be significantly extended compared with supercharging with just one single-stage supercharger unit. Predominantly complex valves for control, for operating-point-dependent cut-in or for defined distribution of the exhaust-gas mass flow are required. The aim here is both to improve the steady-state air supply and to speed up the transient response of the build-up of boost pressure. This in turn improves both the specific consumption and the controllability of the transient operating characteristics and with it the engine emission behavior.

Bi-turbo supercharging

In the case of bi-turbo supercharging, the air supply, instead of using one large exhaust-gas turbocharger, is distributed to two or more smaller exhaust-gas turbochargers connected in parallel. These are generally permanently assigned to specific cylinders on the exhaust-gas side, for example the bank of a V-engine. Each turbocharger pressurizes the engine with air over the full rotational-speed range. The advantage of this arrangement lies primarily in the rapid transient build-up of boost pressure and in the comparatively compact routing of pipework. The turbochargers operating in parallel cannot be controlled independently of each other and should in the process always generate an identical boost pressure.

Figure 11: Sequential supercharging
a) Run-up in one-charger operation,
b) Controlled one-charger operation,
c) Two-charger operation.
1 Exhaust-gas turbocharger 1,
2 Exhaust-gas turbocharger 2,
3 Compressor cut-in valve,
4 Turbine cut-in valve,
5 Transverse pipe, 6 Bypass valve.



Sequential supercharging
Sequential supercharging is predominantly used in ship propulsion systems or generator drives. However, because of the high levels of power it can deliver, this arrangement variant is also used in passenger cars. In the case of passenger-car sequential supercharging, as engine load and speed increase, an additional turbocharger is cut in to the basic supercharging process (Figure 11). Thus, in comparison with mono-supercharging (i.e., a single charger which is geared to the rated output), two supercharging optima are achieved.

The sequential-supercharging system contains, in addition to the turbochargers, valves and sensors which facilitate a smooth switchover from one- to two-charger operation and back from two- to one-charger operation. When the engine is in the lower speed range, only the basic turbocharger is active, supplying all the cylinders with air. When a certain speed is reached, the second turbocharger is run up and cut in. Both turbochargers are now operating in parallel to pressurize all the cylinders. Sharing the air supply between two turbochargers helps to deliver a very wide torque plateau and high rated output. Furthermore, the significantly lower mass moment of inertia of the smaller turbocharger rotor assemblies gives rise to a dynamic response.

An integrated arrangement of both turbochargers including the switchover valves is available on the market, but is complex and only ever specifically suitable. Controlling the switchover process by means of control engineering involves a comparatively complex application of the control parameters.

Sequential supercharging is used in both gasoline and diesel engines [7].

Dual-stage controlled supercharging
In contrast to sequential supercharging, with dual-stage controlled supercharging the turbochargers are connected in such a way as to effect serial operation in certain operating phases (Figure 12). The advantage of this supercharging system over the

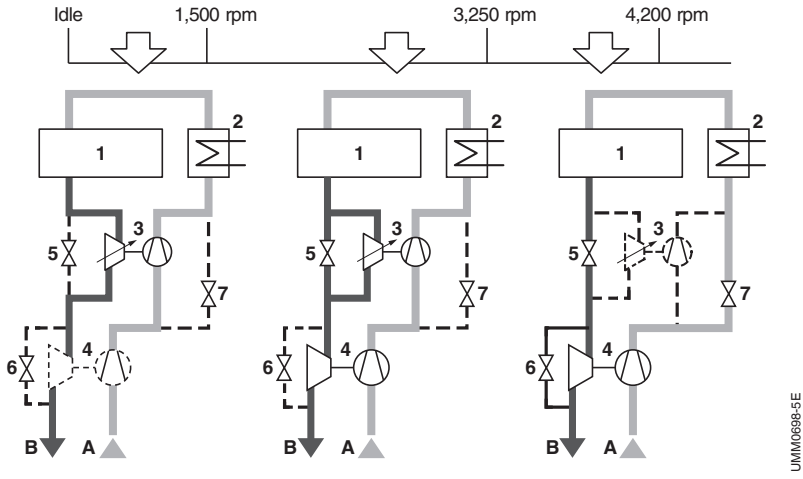
single-stage method lies in the increase in rated output with a simultaneous improvement in the stationary torque at low speeds and in the engine's accelerating performance thanks to the rapid build-up of boost pressure. This extremely powerful system features the use of two turbochargers of significantly different sizes. A dip in torque in the switchover range when this system is appropriately tuned and controlled is largely avoidable.

The entire drawn-in fresh-air mass flow flows through both serial compressor stages for lower full load. As a result of precompression, the relatively small high-pressure compressor works at a higher pressure level so it is able to make available the required boost pressure already for low engine speeds. Performance can be further improved by cooling the charge air after the first compression stage.

Particular importance is attached to the sealing and the control quality of the turbine bypass valve in the high-pressure stage. By way of this valve, depending on the operating point, either the entire exhaust-gas mass flow is expanded in two stages in the high-pressure turbine and then in the low-pressure turbine, or some of the exhaust-gas mass flow is diverted past the high-pressure turbine and delivered directly to the low-pressure turbine. The thermomechanical and vibrational demands placed on the valve system are very tough. This also applies to the rendering of low-loss flow routing between the compressor and turbine stages.

The bypass valves are closed in the engine's low speed range, i.e., with small exhaust-gas mass flow rates. The exhaust-gas mass flow expands primarily through the small high-pressure turbine. This results in a very good transient build-up of boost pressure. As engine speed increases, the high-pressure turbine bypass valve routes an increasing amount of the exhaust-gas mass flow directly to the low-pressure turbine until finally the low-pressure turbine too is relieved by a further wastegate. Thus for small throughputs the charge air is compressed predominantly in the high-pressure stage; however, as throughputs increase it is compressed increasingly in the low-pressure stage. When a specific engine speed is reached, the high-pressure compressor bypass valve is additionally

Figure 12: Schematic diagram of dual-stage controlled supercharging
1 Engine, 2 Intercooler, 3 Turbocharger high-pressure stage,
4 Turbocharger low-pressure stage, 5 Turbine bypass valve,
6 Wastegate, 7 Compressor bypass valve.
A Intake air, B Exhaust gas.



opened and establishes a direct connection from the low-pressure compressor to the intercooler.

In this operating mode dual-stage controlled supercharging enables infinitely variable adaptation of turbocharging to the requirements of engine operation.

The compressor bypass valve can be designed as a (passive) non-return valve or as an actively actuated flap with actuator. With the latter variant the pressure loss in the compressor bypass can be reduced. But this involves increased control-engineering complexity.

Both fixed-geometry turbines and VTG turbines, in each case with an additional bypass valve, are used in the high-pressure stage. Either a wastegate turbine or a variable-geometry turbine (usually without an additional bypass) can be used as the low-pressure turbine.

Dual-stage controlled supercharging was introduced as standard in passenger-car diesel engines in 2004. Currently, the spread of two-stage supercharging systems for use in passenger-car diesel engines is dropping off because they negatively influence the engine emission behavior during the warm-up phase due to their high thermal inertia.

Dual-stage supercharging systems can also be used in gasoline engines and may because of the high boost-pressure requirements be of interest particularly for Miller combustion processes.

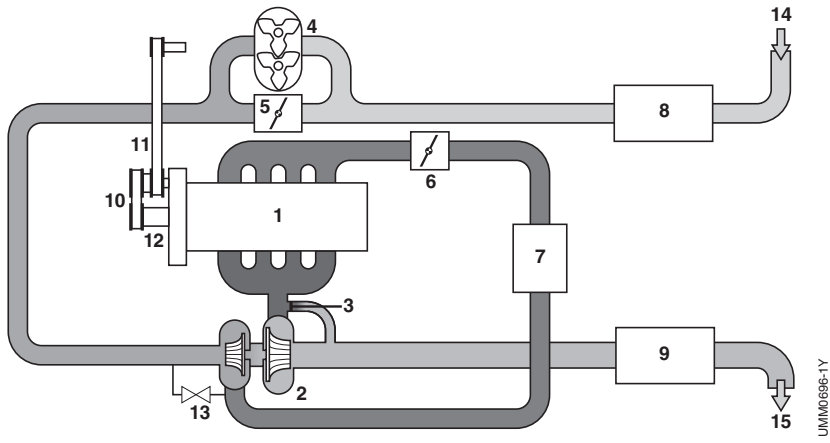
Combination supercharging

Combination supercharging combines a mechanically driven supercharger unit with an exhaust-gas turbocharger (Figure 13). This powerful supercharging system was used for the first time in a passenger car in 1985 (Lancia Delta). It can be used in comparatively small-capacity gasoline engines with high rated output and high torque already at comparatively low speeds and high dynamic response.

In the lower engine-speed range in which little exhaust-gas energy is available to drive the turbocharger turbine, the air supply is practically managed by the mechanically driven supercharger on its own. Because the exhaust-gas turbocharger is designed for the mean-speed and rated-output ranges, this results in a charger with high throughput, optimized for

Figure 13: Combination supercharging

1 Internal-combustion engine, 2 Turbocharger, 3 Wastegate, 4 Compressor (mech. driven supercharger), 5 Control flap, 6 Throttle valve, 7 Intercooler, 8 Air filter, 9 Catalytic converter, 10 Belt drive for accessories with clutch, 11 Belt drive for compressor, 12 Crankshaft, 13 Blow-off valve, 14 Fresh air, 15 Exhaust gas.



rated-output operation. In a certain operating range the system operates as a series connection of two supercharger units, i.e., the total pressure ratio is the product of the individual pressure ratios.

The mechanically driven supercharger can be positioned in the flow direction upstream or downstream of the turbocharger compressor. When the engine is in the upper speed range and there is therefore an adequate supply of exhaust-gas energy, supercharging is performed solely by the turbocharger. The mechanically driven supercharger is bypassed and mechanically isolated as otherwise, even when the mechanically driven supercharger is idling, its friction losses will be removed as drive power from the crankshaft.

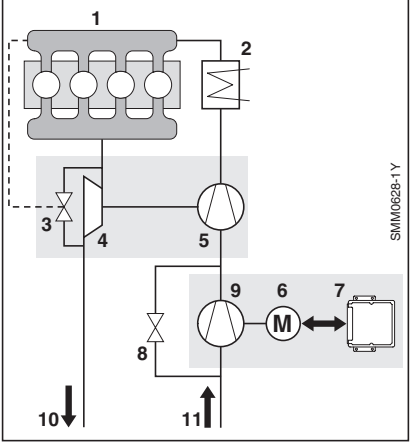
Electrically driven compressor in series with an exhaust-gas turbocharger

One possible solution for obtaining the benefits of exhaust-gas turbocharging in quasi-stationary operation and in the process improving the response of a turbocharger designed for high rated output is the series connection of an exhaust-gas turbocharger with an electrically driven flow compressor ("booster", Figure 14). The advantages of this solution over the electrically assisted exhaust-gas turbocharger are that the effective program-map range is extended by the series connection of two flow compressors and the booster can be placed in thermally suitable locations in the engine compartment. The booster can be positioned in the flow direction before or after the turbocharger. A bypass allows the booster to be bypassed when it is not in operation.

This electrically assisted extension of the engine map is useful for both steady-state and dynamic engine operation. Depending on the energy availability of the vehicle electrical system used, however, different operating strategies must be chosen for a suitable charge balance of the electrical energy accumulator. Depending on the position of the combustion limits for a fault-free, reproducible combustion process, a part-load application is also useful and uncritical in terms of energy. On the basis of the current CO_2 balance and the emission legislation, non-engine-dependent boost-pressure provision is real driving operation is of significant benefit when compared with other applications. The additional function as a secondary-air pump and transient support of cost-sensitive hybrid applications are mentioned as examples.

Figure 14: Electrically driven compressor in series with an exhaust-gas turbocharger

1 Internal-combustion engine, 2 Intercooler,
3 Bypass valve for turbocharger,
4 Turbocharger turbine,
5 Turbocharger compressor,
6 Electric motor,
7 Triggering electronics for electric motor,
8 Bypass valve for eBooster,
9 eBooster compressor, 10 Exhaust gas,
11 Intake air.



For the system to operate effectively, the booster must reach its target pressure ratio within the shortest possible time. When the acceleration of the shaft is taken into consideration, this produces a brief booster power peak. This requires a stabilization of the vehicle electrical system, especially for a vehicle system voltage of 12 to 24 V.

This supercharging system is used both in some luxury-class vehicles with 48 V vehicle electrical systems and in gasoline applications with 12 V vehicle electrical systems.

Current developments are concerned, aside from with engine applications, with the air supply and the recirculation of fuel cells. The assemblies are therefore designed for vehicle system voltages of up to 400 V and a maximum output of 20 kW.

References

[1] H. Hiereth, P. Prenninger: Aufladung der Verbrennungskraftmaschine. Springer-Verlag, 2003.
[2] H. Pucher, K. Zinner: Aufladung von Verbrennungsmotoren: Grundlagen, Berechnungen, Ausführungen. 4th Edition, Verlag Springer Vieweg, 2012.
[3] Otto Cordier, Zeitschrift Brennstoff-Wärme-Kraft, 1953, p. 337ff.
[4] D. Wolf: Das CORDIER-Diagramm unter besonderer Berücksichtigung der axialen Turboarbeitsmaschine. Dissertation, 2009.
[5] H. Sigloch: Strömungsmaschinen, Grundlagen und Anwendungen. 6th Edition, Hanser-Verlag, 2018.
[6] R. v. Basshuysen (Editor): Ottomotor mit Direkteinspritzung. 3rd Edition, Verlag Springer Vieweg, 2013.
[7] J. Portalier, J.C. Blanc, F. Garnier, N. Schorn, H. Kindl: Twin Turbo Boosting System Design for the New Generation of PSA 2.2 liter HDI Diesel Engine. 11th Supercharging Conference, Dresden, 2006.
[8] M. Mayer: Abgasturbolader – Sinnvolle Nutzung der Abgasenergie. 6th Edition, Verlag Moderne Industrie, 2011.
[9] G. Hack, G.-I. Langkabel: Turbo- und Kompressormotoren – Entwicklung und Technik. 3rd Edition, Motorbuch-Verlag, 2003.

Exhaust-gas system

Purpose and design

In compliance with legal requirements, the exhaust-gas system reduces the pollutants in the exhaust gas that are generated by an internal-combustion engine. The exhaust-gas system also helps to muffle exhaust-gas noise and to discharge the exhaust gas at a convenient point on the vehicle. Engine power should be reduced as little as possible during the process.

Components
An exhaust-gas system consists of the exhaust manifold, the components for exhaust-gas treatment and for sound absorption, and the connections between these components.

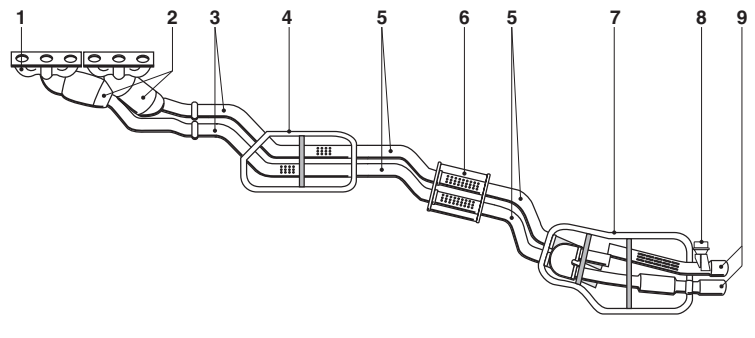
The design and the layout of these components are very different for passenger cars and commercial vehicles. In the case of passenger-car exhaust-gas systems the components are usually individually connected by pipes and routed under the vehicle chassis (Figure 1). Depending on engine displacement and the type of muffler used, a passenger-car exhaust-gas system weighs between 8 and 40 kg. The components are primarily made of high-alloy steels, as they are exposed to corrosion attacks from the inside by hot gas and condensate, and from the outside by moisture and splash and salt water.

Since the introduction of Euro IV, commercial vehicles have also been required to carry components for exhaust-gas treatment. These are, however, usually integrated into a large system and secured to the frame (see Commercial-vehicle exhaust-gas systems). The individual components are explained in the following using a passenger-car exhaust-gas system as the example

Exhaust-gas treatment
The components of the exhaust-gas treatment include
– the catalytic converter to break down the gaseous pollutants in the exhaust gas, and
– the particulate filter (or soot filter) to filter out the fine, solid particulates in the exhaust gas (especially in diesel engines).

Catalytic converters are installed in the exhaust-gas system as close as possible to the engine so that they can quickly reach their operating temperature and therefore be effective even at low operating temperatures (e.g. in urban driving). The decisive factor here is the light-off temperature of the catalytic converter, i.e. the temperature at which the catalytic converter begins to break down pollutants (approx. 250 °C for three-way convert-

Figure 1: Exhaust-gas system of a passenger car (example with three mufflers)
1 Exhaust manifold, 2 Near-engine catalytic converter, 3 Front pipe, 4 Front muffler, 5 Intermediate pipe, 6 Center muffler, 7 Rear muffler, 8 Exhaust-gas flap, 9 Tailpipe.



ers). Some catalytic-converter coatings, such as those for NO_x storage converters for example, are highly sensitive to temperature. These catalytic converters are therefore sooner installed in the cooler underbody area.

Diesel particulate filters are also installed in the front area of the exhaust-gas system to ensure that the soot particles they have retained are burnt off more effectively at the higher exhaust-gas temperatures.

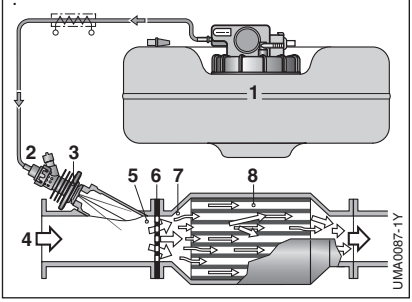
Essentially three-way catalytic converters are used in spark-ignition engines. During engine operation with a stoichiometric air-fuel mixture ($\lambda = 1$) they convert the hydrocarbons (HC), nitrogen oxides (NO_x) and carbon monoxide (CO) present in the untreated exhaust gas by more than 99 %. An additional NO_x storage catalytic converter is only installed in lean-burn-operated gasoline direct-injection engines. This reduces the nitrogen oxides (NO_x) increasingly produced in this operating state.

Diesel engines require an oxidation-type catalytic converter to oxidize the hydrocarbons (HC) and carbon monoxide (CO) and a particulate filter to hold back the solid exhaust-gas components. For low NO_x emissions a further catalytic converter for NO_x reduction is required – an NO_x storage catalytic converter or an SCR catalytic converter (Selective Catalytic Reduction).

In order to ensure that the emissions are converted as much as possible in the catalytic converter or are well filtered in the particulate filter, it is essential to optimize the flow against these components in the exhaust-gas system. This is generally achieved by the shaping of the inflow funnel. Additional components such as swirl or mixing elements are required to optimize uniform distribution further. Especially SCR systems in which a reducing agent (Urea/water solution) is injected in liquid form into the exhaust-gas system require a mixer in order to ensure vaporization or atomization of the reducing agent to NH_3. The aim is to ensure uniform and gaseous distribution of NH_3 ahead of the catalytic converter (Figure 2).

Figure 2: Atomization and vaporization of reducing agent in SCR system

1 Reservoir for reducing agent,
2 Connection, injector for reducing agent,
3 Injector for reducing agent,
4 Gas flow, 5 Exhaust pipe, 6 Mixer,
7 Inflow funnel, 8 SCR catalytic converter.



Sound absorption

The primary cause of exhaust-gas noise is the gas pulsations of the internal-combustion engine, i.e. the gas vibrations which are generated by the combustion process and by the exhaust gas being forced out through the exhaust valves during each engine power cycle. This pulsation noise is in fact damped slightly by the catalytic converters and particulate filter. But this is not sufficient to keep the noise below the pass-by limit value prescribed by law (see Vehicle acoustics). For this reason, mufflers are installed in the middle or rear section of the exhaust-gas system.

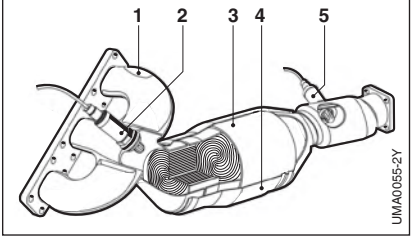
Depending on the number of cylinders and engine output, generally one, two or three mufflers are used in an exhaust-gas system. In V-engines, the left and right cylinder banks are often run separately, each being fitted with its own catalytic converters and mufflers.

The noise-emission limit for the complete vehicle is defined by legislation. The noise produced by the exhaust-gas system represents a substantial source of noise emission in a vehicle. This fact makes it necessary to devote particular attention and resources to the development of mufflers. Although the aim is to reduce noise in compliance with the legislation, they can also create the sound specific to the type of vehicle (sound design).

Exhaust manifold

An important component in the exhaust-gas system is the exhaust manifold (Figure 3). It routes the exhaust gas out of the cylinder outlet ports into the exhaust-gas system. The geometric design of the manifold (i.e. length and cross-section of the individual pipes) has an impact on the performance characteristics, the acoustic behavior of the exhaust-gas system, and the exhaust-gas temperature. In some cases, the exhaust manifold is air-gap-insulated (i.e. double-walled) in order to achieve high exhaust-gas temperatures quickly so as to enable the catalytic converter to work at an early stage during cold start. The exhaust manifold must consistently endure very high temperatures (up to 1,050 °C in gasoline engines). It is therefore produced from very high-quality material (high-alloy casting or high-alloy stainless steel).

Figure 3: Exhaust manifold with near-engine catalytic converter
1 Exhaust manifold, 2 λ sensor, 3 Metal monolith, 4 Insulating shell, 5 OBD sensor.



Catalytic converter

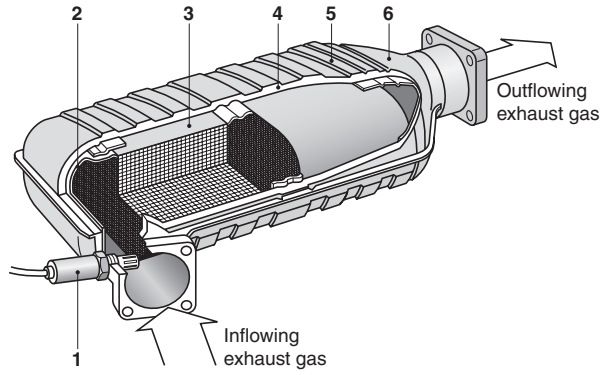
A catalytic converter is made up of an inflow funnel, an outflow funnel and a monolith (Figure 4). The monolith contains a large number of very fine, parallel channels covered with an active catalytic coating. The number of channels ranges from 400 to 1,200 cpsi (cells per square inch). The functional principle of the active catalytic-converter layer is described elsewhere (see Catalytic exhaust-gas treatment). For operational reasons, several differently coated monoliths are often used in one catalytic converter. Particular attention must be paid to the shape of the inflow funnel to ensure that the exhaust gas flows uniformly through the monolith. The external shape of the monolith depends on the installation space available in the vehicle, and may be triangular, oval or round.

The monolith can be made of metal or ceramic material.

Metal monolith

The metal monolith is made of finely corrugated, 0.05 mm thick metal foil, wound and brazed in a high-temperature process. Due to the very thin walls between the channels, the metal monolith offers an extremely low resistance to the exhaust gas. It is therefore frequently used in high-performance vehicles. The metal monolith can be welded directly to the funnels.

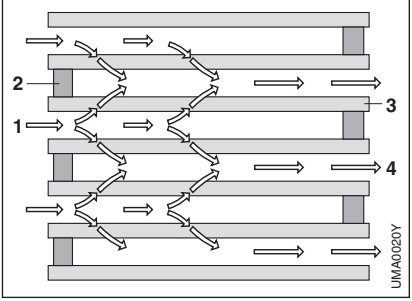
Figure 4: Catalytic converter with ceramic monolith
1 λ sensor,
2 Inflow funnel,
3 Ceramic monolith,
4 Mounting mat,
5 Metal housing,
6 Outflow funnel.



Ceramic monolith
The ceramic monolith is made of cordierite. Depending on the cell density, the wall thickness between the channels ranges from 0.05 mm (at 1,200 cpsi) to 0.16 mm (at 400 cpsi).

Ceramic monoliths have an extremely high stability to temperature and thermal shock. However, they cannot be installed directly in the metal housing and require a special mounting. This mounting is necessary in order to compensate for the difference between the thermal-expansion coefficients of steel and ceramics, and to protect the sensitive monolith against shocks. Extreme care and attention is required in the mounting and production process, particularly for thin-walled monoliths (< 0.08 mm). The monolith is mounted on a mat located between the metal housing and the ceramic monolith. The mounting mat is made up of ceramic fibers. It is extremely flexible to ensure that the pressure load exerted on the monoliths is at a minimum. It also serves as a heat insulator.

Figure 5: Ceramic particulate filter
1 Exhaust-gas inlet,
2 Ceramic plug,
3 Cell partition,
4 Exhaust-gas outlet.



Particulate filter

As with catalytic-converter monoliths, there are metallic and ceramic particulate-filter systems. The method of installing and mounting this filter in the metal housing is the same as the process used for the catalytic converter.

Design
In the same way as the ceramic monolith for the catalytic converter, the ceramic particulate filter is made up of a large number of parallel channels. However, these channels are alternately closed (Figure 5). Consequently, the exhaust gas is forced to flow through the porous walls of the honeycomb structure. The solid particles are deposited in the pores. Depending on the porosity of the ceramic body, the filtration efficiency of these filters can attain up to 97%.

Regeneration
The soot deposits in the particulate filter induce a steady rise in flow resistance. For this reason, the particulate filter must be regenerated at certain intervals in two different processes. Details on these processes are described elsewhere (see Exhaust-gas treatment for diesel engines).

Passive process
In the passive process, the soot is burnt off by a catalytic reaction. For this purpose, an additive in the diesel fuel reduces the flammability of the soot particles to normal exhaust-gas temperatures.

Other passive regeneration options include catalytically coated particulate filters or the CRT process (see Continuous Regeneration Trap).

Active process
In the active process, external measures are implemented to heat the filter to the temperature necessary for burning off the soot. This rise in temperature can be achieved by a burner mounted upstream of the filter or by secondary injection initiated by the engine management and the use of a preliminary catalytic converter.

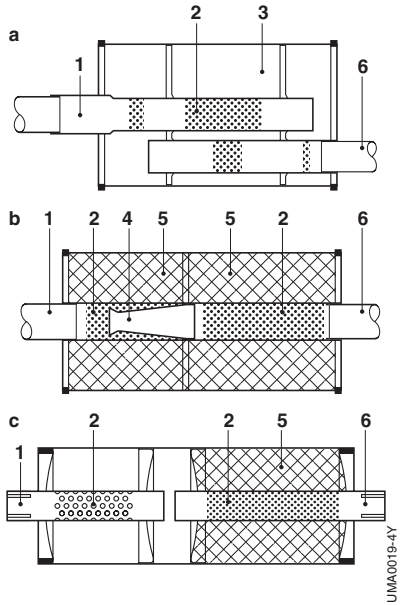
Mufflers

Mufflers (or silencers) are intended to smooth out exhaust-gas pulsations and make them as inaudible as possible. There are essentially two physical principles involved: reflection and absorption. Mufflers also differ according to these principles. However, they mostly comprise a combination of reflection and absorption.

In a reflection muffler (Figure 6a) the sound waves travel through the pipes and are reflected at the pipe outlet in the chambers. In an absorption muffler (Figure 6b) the sound waves spread over the perforation in the absorption material, where they are then muffled. Figure 6c shows a combination of reflection muffler and absorption muffler.

As mufflers and the exhaust-gas system pipes together form an oscillating system with its own natural resonance, the position of the mufflers is highly significant for the quality of sound-damping. The objective is to tune the exhaust-gas systems as low as possible, so that their natural frequencies do not excite bodywork resonances. To avoid structure-borne noise and to provide heat insulation for the vehicle underbody, mufflers often have double walls and an insulating layer.

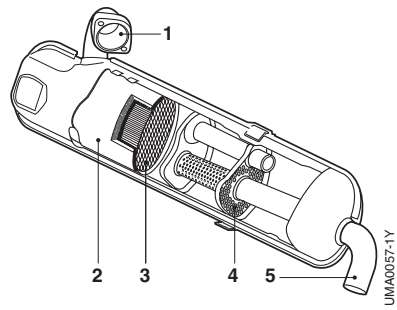
Figure 6: Muffler principles
a) Reflection muffler,
b) Absorption muffler,
c) Combination of reflection and absorption mufflers.
1 Inlet pipe,
2 Perforated pipe,
3 Chamber,
4 Venturi nozzle,
5 Absorption material,
6 Outlet pipe.



Reflection mufflers

Reflection (resonator-type) mufflers consist of chambers of varying lengths which are interconnected by pipes (Figure 6a and Figure 7). The pipes and partitions are perforated, thereby allowing exhaust gases to pass through. The differences in the cross-sections between the pipes and the chambers, the diversion of the exhaust gases, and the resonators formed by the connecting pipes and the chambers generate superpositions of the sound waves moving back and forth, causing the sound waves to be partially extinguished. In this way, effective muffling is achieved, particularly in the middle and low frequency ranges. The more such chambers are used, the more efficient is the muffler.

Figure 7: Muffler with integrated catalytic converter
1 Inlet pipe,
2 Mounting mat,
3 Ceramic monolith,
4 Reflection muffler,
5 Tailpipe.



Absorption mufflers

Absorption mufflers are designed with one chamber, through which a perforated pipe passes (Figure 6b). The chamber is filled with absorption material (mineral wool or glass fibers). The sound enters the absorption material through the perforated pipe and is converted into heat by friction.

The absorption material usually consists of long-fiber mineral wool or glass fibers with a bulk density of 100 to 150 g/*l*. The level of muffling depends on the bulk density, the sound-absorption grade of the material, and on the length and coating thickness of the chamber. Damping takes place across a very broad frequency band, but only begins at higher frequencies.

The shape of the perforations, and the fact that the pipe passes through the wool, ensures that the material is not blown out by exhaust-gas pulses. Sometimes the mineral wool is protected by a layer of stainless-steel wool around the perforated pipe.

Because the exhaust gas in absorption mufflers essentially passes through a straight pipe, the pressure loss here is very low compared with the pressure loss at the differences in the cross-sections in reflection mufflers.

Muffler design

Depending on the space available under the vehicle, mufflers are produced either as spiral-wound casing or from half-shells.

To produce the jacket for a spiral-wound muffler, one or several metal sheet blanks are shaped over a round mandril and joined together either by the longitudinal folds or by laser welding. The completely assembled and welded core is then installed in the jacket casing. It consists of internal tubes, baffles, and intermediate layers. The outer layers are then connected to the jacket in a folding or laser-welding process.

It is often not possible to effectively accommodate a spiral-wound muffler in view of the complicated space conditions in the floor assembly. In such cases, a shell-type muffler made of deep-drawn half-shells is used as it can assume virtually any required shape.

The total volume of the mufflers in a passenger-car exhaust-gas system corresponds to approximately eight to twelve times the engine displacement.

Connecting elements

Pipes are used to connect the catalytic converters and mufflers together. Arrangements where the catalytic converter and muffler are integrated in a single housing may also be used on very small engines and vehicles (Figure 7).

The pipes, the catalytic converter and muffler are connected to form an integrated system by means of plug-in connections and flanges. Many original-equipment systems are fully welded for faster mounting.

The entire exhaust-gas system is connected to the vehicle underbody via elastic mounting elements (Figure 8). The fixing points must be carefully selected, otherwise vibration may be transmitted to the bodywork and generate noise in the passenger compartment. The wrong fixing points may also create strength and therefore durability problems. In some cases, these problems are counteracted by the use of vibration absorbers. These components oscillate at the critical frequency in precisely the opposite direction of the exhaust-gas system, thereby absorbing system vibration.

Figure 8: Mounting element
1 Rubber mount,
2 Metal bracket,
3 Muffler shell.

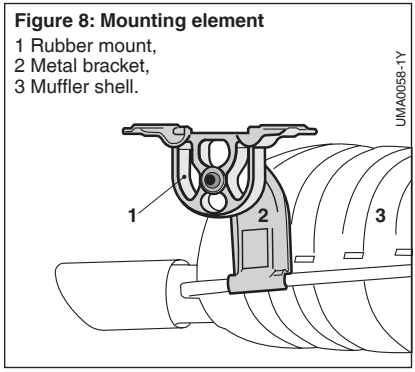
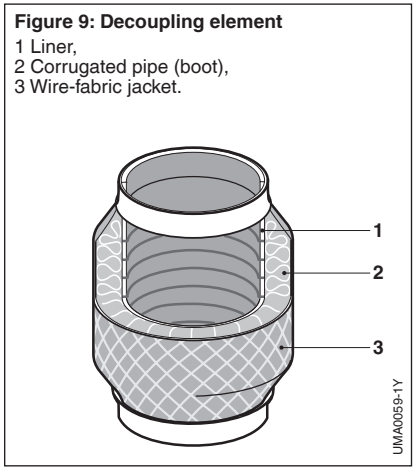


Figure 9: Decoupling element
1 Liner,
2 Corrugated pipe (boot),
3 Wire-fabric jacket.



The exhaust-gas system noise at the exhaust-emission point (tailpipe), as well as sound radiation from the mufflers, can also cause bodywork resonance. Depending on the intensity of the engine vibrations, decoupling elements (Figure 9) are used to isolate the exhaust-gas system from the engine block and to relieve the stress load on the exhaust-gas system. A decoupling element comprises the liner, i.e. an inner pipe with pipe segments that move in each other. The corrugated pipe, also known as the boot, is mounted on top. The liner serves to protect the boot and limit linear expansion. The corrugated pipe exercises the decoupling/isolating effect by way of its soft structure. The corrugated pipe is protected against external influences by a wire-fabric jacket.

Finally, the mounting arrangement of an exhaust-gas system is tuned so that it is rigid enough to withstand vibrations reliably on the one hand, and it exhibits sufficient flexibility and damping properties to reduce the transfer of forces to the bodywork effectively on the other.

Acoustic tuning devices

A number of different components can be used to eliminate disturbing frequencies in the noise emitted from the tailpipe. These components can be used specifically to damp individual frequency ranges very effectively.

Helmholtz resonator
The Helmholtz resonator consists of a pipe arranged along the side of the exhaust-gas train and a defined volume connected to it (Figure 10). The gas volume acts as a spring, while the gas in the pipe section acts as a mass. At its resonant frequency, this spring-mass system provides a very high degree of sound absorption but in a narrow frequency band. The resonant frequency f is dependent on the size of the volume V as well as the length L and the cross-sectional area A of the pipe:

$$f = \frac{c}{2\pi} \sqrt{\frac{A}{L \cdot V}}$$

The value c is the speed of sound.

λ/4 resonators
λ/4 resonators consist of a pipe, sealed at the end, branching off from the exhaust-gas system. The resonant frequency f of these resonators is derived from the length L of the pipe branch. It is expressed as:

$$f = \frac{c}{4L} .$$

These resonators also feature a very narrowband damping range about their resonant frequency.

Exhaust-gas flaps
Exhaust-gas flaps are most commonly found in rear mufflers. Depending on the engine speed or exhaust-gas throughput, they close off a bypass pipe in the muffler or a second tailpipe (Figure 11). As a result, exhaust-gas noise can be substantially damped at lower engine speeds without the need to trade off power losses at high engine speeds.

Exhaust-gas flaps can be either self-controlling based on pressure and flow, or they can be controlled externally. An interface to the engine management system must be provided for externally controlled flaps. This makes them more complex than self-controlled flaps. However, their application range is also more flexible.

Figure 10: Helmholtz resonator
1 Helmholtz volume,
2 Gas-carrying exhaust pipe,
3 Helmholtz pipe,
4 Muffler shell.

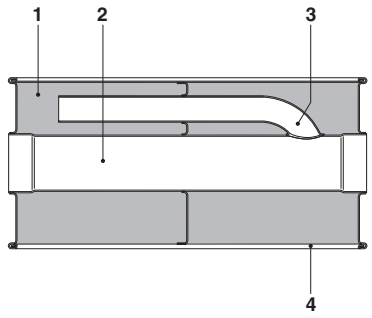
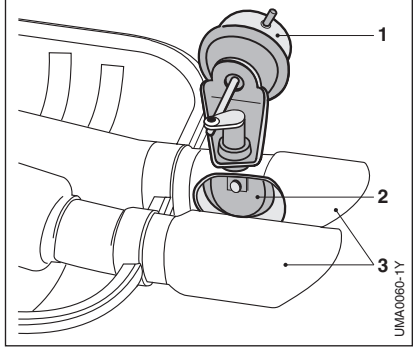


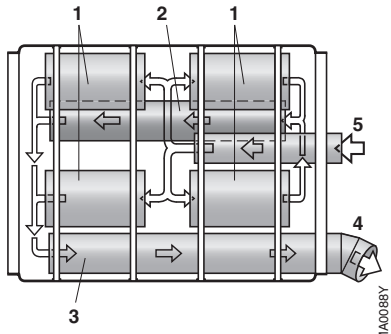
Figure 11: Exhaust-gas flap, controlled by vacuum
1 Vacuum unit,
2 Butterfly valve,
3 Tailpipes.



Commercial-vehicle exhaust-gas systems

In commercial-vehicle exhaust-gas systems most of the previously described components are integrated in a housing which is mounted to the vehicle frame. The number of catalytic converters and particulate filters depends on which emissions-control legislation the exhaust-gas system is designed for.

Figure 12: Commercial-vehicle exhaust-gas system for Euro IV
1 SCR catalytic converters, 2 Inlet pipe,
3 Outlet pipe, 4 Tailpipe, 5 Exhaust-gas inlet.

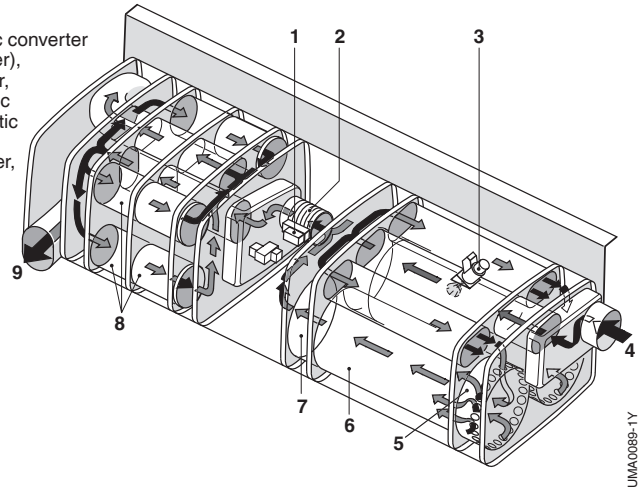


Exhaust-gas systems for Euro IV and Euro V

Generally speaking, a particulate filter is not required for Euro IV and Euro V. Only oxidation and SCR catalytic converters are used (see Exhaust-gas treatment for diesel engines). Alternatively, the diesel engine can also be tuned in such a way that the untreated NO_x emissions comply with the limit values for Euro IV and Euro V. But then a particulate filter is required in the exhaust-gas system.

Figure 12 shows a Euro IV exhaust-gas system with SCR catalytic converters. In contrast to passenger-car exhaust-gas systems, several catalytic converters are frequently arranged in parallel in order to accommodate the required catalytic surface in the space available. Bypass pipes and holes in the bases serve to route the gases and muffle the sound. Depending on the engine size, these systems attain a volume of 150 to 200 *l* and a weight of 150 kg.

Figure 13: Commercial-vehicle exhaust-gas system for Euro VI and EPA 10
1 Urea injector,
2 Decoupling element,
3 HC injector,
4 Inlet pipe,
5 Pre-oxidation catalytic converter (pre catalytic converter),
6 Diesel particulate filter,
7 Post-oxidation catalytic converter (post catalytic converter),
8 SCR catalytic converter,
9 Outlet pipe.



Exhaust-gas systems for EuroVI and EPA10

Exhaust-gas systems for the latest legislation (Euro IV in Europe and EPA 10 in the USA) require all the components, i.e. oxidation catalytic converters, particulate filters and SCR catalytic converters (Figure 13). They are thus even bigger in terms of volume and weight.

Two concepts are currently used. Either all the components are accommodated in a single housing or the SCR catalytic converters and the particulate filters are divided between two housings. The following components are additionally required to ensure correct functioning of exhaust-gas treatment: The SCR function requires a urea metering system (see SCR system), whose nozzle (injector) is located at a suitable location in the exhaust-gas system. Furthermore, to ensure reliable regeneration of the particulate filter, an HC metering unit is frequently needed to inject fuel (see HCI system). In both cases a position must be chosen which ensures that the liquid urea solution and the fuel are adequately vaporized and distributed. If necessary, mixers are used here to prepare the liquids better (see section on Exhaust-gas treatment).

A variety of sensors must also be accommodated in the housings. In addition to pressure sensors for monitoring the filter load, temperature sensors and NO_x sensors for monitoring NO_x conversion are required. The position of the sensors in the exhaust-gas system must, depending on the system design, be optimized to guarantee an adequate signal quality in all operating states.

References

[1] C. Hagelüken: Autoabgaskatalysatoren – Grundlagen, Herstellung, Entwicklung, Recycling, Ökologie. 3rd Edition, Expert-Verlag, 2016.

Management for spark-ignition engines

Description of the engine management system

The engine management system ensures that the driver command is implemented. The driver can request acceleration, deceleration or driving at a constant speed – the engine management system ensures that the spark-ignition engine's drive output required to achieve this is set. The system controls all engine functions in such a way that the engine delivers the required level of torque, but fuel consumption and exhaust emissions are kept low.

The power output from a spark-ignition (SI) engine is determined by the available clutch torque and the engine speed. The clutch torque is the torque developed by the combustion process less friction torque (friction losses in the engine), pumping losses, and the torque needed to drive the auxiliary equipment (Figure 1). The drive torque is available at the wheels. This is developed from the clutch torque reduced by the losses in the clutch and in the transmission. This resulting torque is offset by the tractive resistances such as rolling resistance and aerodynamic drag. Depending on the driver command, a state of equilibrium or imbalance exists between these resistances and the drive torque. When the resistances and the drive torque are equal, a constant driving speed is adopted, otherwise the vehicle is subject to acceleration or deceleration.

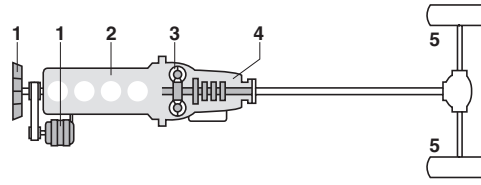
Combustion torque is generated in the engine's power cycle and is mainly determined by the following variables:
– The air mass which is available for combustion when the intake valves close,
– The fuel mass available in the cylinder,
– The point at which the combustion takes place.

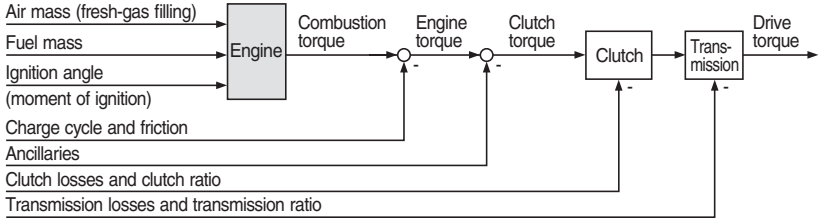
Furthermore, there are also minor influences, for example, due to mixture composition (amount of residual exhaust gas) or combustion processes.

The primary function of engine management is to coordinate the various subsystems (air, fuel and ignition systems) in order to adjust the torque requested by the engine and at the same time meet the exacting standards placed on exhaust-gas emissions, fuel consumption, power output, comfort, and safety. The engine-management system also performs diagnoses on the subsystems.

Figure 1: Drivetrain torques

1 Auxiliary systems (alternator, A/C compressor, etc.),
2 Engine,
3 Clutch,
4 Transmission,
5 Drive wheels.





System overview

Electrical system overview

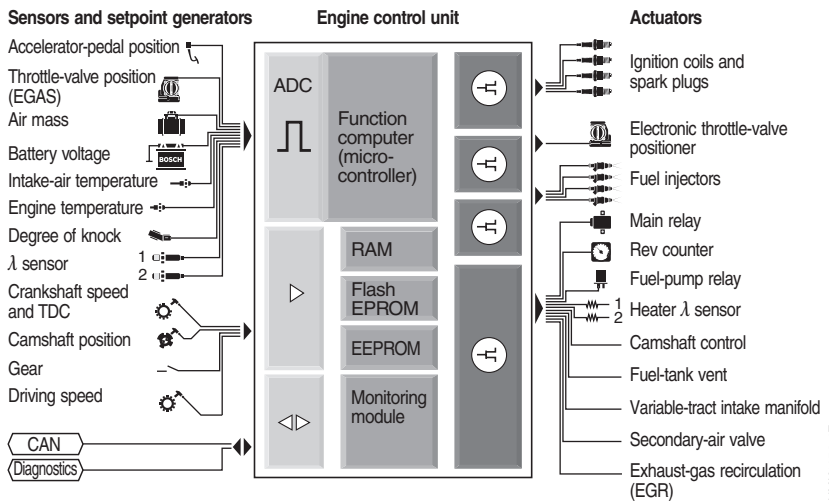
Motronic is the name given to Bosch systems for the open- and closed-loop control of spark-ignition (SI) engines. The Motronic system (Figure 2) comprises all the sensors for detecting ongoing operating data from the engine and vehicle, and all the actuators that perform the required adjustments to the SI engine. The control unit uses data from the sensors to scan the status of the vehicle and engine at very short intervals (every few milliseconds to meet the system's real-time requirements). Input circuits suppress sensor-signal interference and convert the signals to a single uniform voltage scale. An analog/digital converter then transforms the conditioned analog signals into digital values. Other signals are received via digital interfaces (e.g. CAN bus, FlexRay) or via pulse-width-modulated (PWM) interfaces.

The nerve center of the engine control unit is a microcontroller with a program memory (e.g. flash EPROM), which stores all the process control algorithms – i.e. arithmetical processes performed according to a specific pattern – and data (parameters, characteristics, program maps). The input variables derived from the sensor signals influence the calculations in the algorithms, and thus the triggering signals for the actuators. From these input signals, the microcontroller detects the sort of vehicle response the driver wants and calculates, for example, the required torque, the resulting cylinder charge with the associated injected fuel quantity, the correct ignition timing, and the triggering signals for the actuators (e.g. the evaporative-emissions control system, the exhaust-gas turbocharger, and the secondary-air system).

The low-level signal data from the microcontroller outputs are adapted by the driver stages to the levels required by the various actuators.

Another important function of Motronic is the monitoring of the operational ability of the entire system, using On-Board Diagnostics (OBD). Legal requirements (diagnosis legislation) have placed extra demands on Motronic, with the result that about half of Motronic system capacity (in terms of computing power and memory requirements) is dedicated to diagnostics-related tasks.

Figure 2: Components used for electronic control of a spark-ignition engine



Functional system overview
An engine management system has many other secondary functions in addition to its primary functions of cylinder-charge control, fuel supply, mixture formation, and ignition. The entire system is broken down into subsystems so that they can be described more clearly and in greater detail. The full description is shown in the system structure (Figure 3).

Torque Demand (TD) subsystem
The driver issues a concrete driver command through the accelerator-pedal position. The accelerator-pedal position is converted into a setpoint value for the drive torque.

As well as a direct torque input, the driver can also input an indirect driver command via the cruise-control system. A setpoint drive torque is calculated, depending on the current driving condition.

When the accelerator pedal is not pressed, that amount of engine torque which is needed to maintain idle speed is calculated.

Drivability filters, the surge-damping function, the electrical system (starter, alternator, battery) and other electrical consumers such as the air-conditioning system issue further torque demands.

Torque Structure (TS) subsystem
The varied torque demands from the Torque Demand subsystem, demands from the transmission system, driving-dynamics demands, and other engine-specific torque demands (e.g. catalytic-converter heating) are coordinated in the Torque Structure subsystem. The result is the generation of a torque demand made of the internal-combustion engine.

Setpoint values for charge, injection and ignition are generated from the resulting torque demand of the engine.

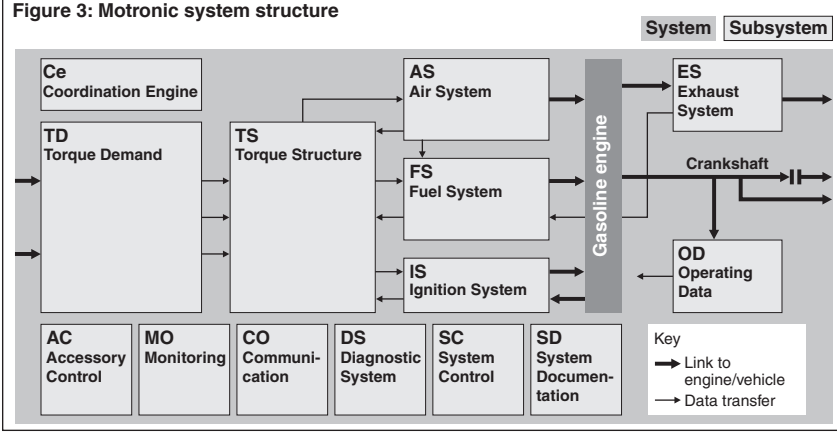
The charge is input as a relative air mass. The relative air mass (standardization for all engine-performance classes) is the ratio between the real air mass in the cylinder and the maximum possible air mass in the cylinder at the current engine speed.

The setpoint value for ignition is described by an ignition angle.

Torque reductions can be represented with injection blank-outs (e.g. traction control system demand). The number of blank-outs is determined for this purpose.

In systems with gasoline direct injection, it is possible to set lean-burn operating modes (e.g. stratified fresh charge in the combustion chamber). In these operating modes, it is also possible to set the engine torque by inputting a λ setpoint value.

Figure 3: Motronic system structure



Physical models calculate the actual engine torque at the clutch from various sensor signals. The actual torque is used to monitor the Motronic system and is required by other systems such as transmission control.

Air System (AS) subsystem
The input of the relative setpoint air mass from the Torque Structure subsystem is converted into concrete quantities for the actuators which are used for cylinder-charge control.

The throttle valve is the main actuator for charge. Models are used to calculate the throttle-valve opening angle from the setpoint air mass and from this opening angle, a pulse-width-modulated activation of the actuator.

There are systems in which the main setting path is represented by the activation of intake and exhaust valves. The throttle valve usually remains constantly open in such systems. Only in special situations (e.g. in limp-home mode) is the throttle valve used in these systems as the setting path for charge.

In turbocharged engines, wastegate activation for the exhaust-gas turbocharger or control of mechanical superchargers are also taken into consideration.

The camshaft adjustment systems and exhaust-gas recirculation valves are further setting devices.

In addition, the current actual charge of the internal-combustion engine is determined. Sensor signals such as pressure and temperature in the intake manifold are used as basic variables for this purpose.

Fuel System (FS) subsystem
The function of the fuel system is to supply the fuel rail with fuel from the fuel tank in the requested quantity and at the prespecified pressure.

By using the current actual charge, the fuel pressure in the rail and the pressure in the intake manifold, the opening duration of the fuel injectors is calculated from the λ setpoint value.

The fuel injectors are activated synchronously with the crank angle in order to optimize the air/fuel mixture.

Longer-term adaptations of the actual value for λ ensure greater pilot-control accuracy of fuel metering.

Ignition System (IS) subsystem
The resulting ignition angle is calculated from the setpoint input for ignition, from the engine operating conditions and on the basis of interventions (e.g. knock control), and an ignition spark is generated at the spark plug at the desired moment of ignition.

The ignition angle is set in such a way that the engine runs with optimal consumption. The subsystem only deviates from this in a few special situations (e.g. catalytic-converter heating or fast torque reduction when changing gear).

The knock-control system permanently monitors combustion in all the cylinders. It ensures that the engine runs with optimal consumption close to the knock limit. At the same time, damage caused by combustion knock is safely avoided. Failures in the knock-detection path are subject to continuous monitoring so that ignition can be effected in the event of a fault at sufficient distance from the knock limit.

Exhaust System (ES) subsystem
The open- and closed-loop control interventions for optimum operation of the three-way catalytic converter are calculated in this subsystem. The combustion mixture must be regulated within narrow limits around the stoichiometric mixture ratio.

The operational capability of the catalytic converter is also monitored. The signals from exhaust-gas sensors (e.g. λ oxygen sensor) serve as the basis for this monitoring.

Component protection functions ensure that the exhaust system is not subject to thermal overloads. The actual temperatures in the exhaust system required for this purpose are usually modeled.

In lean-burn operating modes with stratified fresh charge (in the case of gasoline direct injection), the combustion mixture is also controlled to ensure optimum operation of the NO_x storage catalytic converter.

Coordination Engine (CE) subsystem
In the case of gasoline direct injection, the operating modes (e.g. operation with homogeneous or stratified mixture distribution in the combustion chamber) are coordinated and switched over. In order to determine the required operating mode, the demands of various functionalities must be coordinated on the basis of defined priorities.

Operating Data (OD) subsystem
The Operating Data subsystem evaluates engine operating-condition variables (e.g. engine speed, temperatures), ensures digital conditioning and plausibilization, and makes the result available to the other subsystems.

Adaptation of tolerances in speed sensing helps to control injection and ignition more accurately.

Engine misfires – as a precondition for catalytic-converter protective functions – are detected.

Accessory Control (AC) subsystem
Additional functionalities such as A/C-compressor control, fan control or engine-temperature regulation are frequently integrated into the engine management system. They are coordinated in the Accessory Control subsystem.

Communication (CO) subsystem
The vehicle network contains many other systems (e.g. transmission control, Electronic Stability Program), as well as the Motronic system. The systems exchange information via standardized interfaces (e.g. CAN communication).

In addition, workshop testers can read out signals from the engine management system and perform defined actuator adjustments (actuator diagnosis).

Diagnostic System (DS) subsystem
The operational capability of the Motronic system is subject to continuous monitoring by diagnostic functions. These include both electrical and plausibility checks though comparisons of sensor signals with models. Faults are stored and managed (e.g. faults are assigned "time stamps"). The faults can be called up at a later stage on workshop testers.

Some diagnostic functions are only operational under certain boundary conditions (e.g. in specific temperature or load ranges). There are also diagnostic function which must occur in a prespecified sequence. Coordination of this sequence control is performed by the Diagnostic System.

Monitoring (MO) subsystem
"Drive by wire" systems are relevant to monitoring. The core function is torque comparison. This compares the permissible torque calculated on the basis of driver command with the actual torque calculated from the engine data.

On further levels, the computer core and its peripherals are monitored.

System Control (SC) subsystem
The run-up of the Motronic system is regulated. Computing frameworks must be provided before individual functionalities are calculated. Different computing frameworks (e.g. frameworks with angle or time synchronism) are required for the resource optimization of the computing time.

Defined functionalities (e.g. function diagnosis of output stages) occur before the engine starts. Sequence control also handles resets and ECU run-on.

System Document (SD) subsystem
In addition to the extensive open- and closed-loop control functions of the Motronic system, numerous documents are required in order to describe a concrete project in detail. These documents include system descriptions of the ECU hardware and software, the wiring harness, the engine data, components, and connection-pin assignments.

Motronic versions for passenger cars

History
Originally, Motronic's function essentially combined electronic fuel injection and electronic ignition in a single control unit. Gradually, more and more functions were added in response to legislative requirements to reduce exhaust-gas emissions, reduce fuel consumption, and meet increased demands in terms of performance, driving comfort, and driving safety. Examples of these additional functions are:
– Idle-speed control,
– λ control,
– Control of the evaporative-emissions control system,
– Exhaust-gas recirculation to reduce NO_x emissions and fuel consumption,
– Control of the secondary-air system to reduce HC emissions in the starting and warm-up phases,
– Control of the exhaust-gas turbocharger and variable-tract intake manifold to increase engine performance,
– Camshaft control to reduce exhaust-gas emissions and fuel consumption and to increase performance,
– Component protection (e.g. knock control, engine-speed limitation, exhaust-gas temperature control).

The basis for the first Motronic, which was introduced from 1979 onwards in production vehicles, was the L-Jetronic fuel-injection system with air-flow sensing and electronic ignition with rotating distributor system for a manifold-injection engine with individual-cylinder injection and a conventional, mechanical throttle valve. This M-Motronic was further developed like the injection and ignition systems further used in parallel for this purpose. For load sensing the air-flow sensor was replaced by the air-mass meter, which enables the engine load to be sensed more accurately. Electronic ignition with rotating ignition distributor system was replaced by fully electronic ignition with static distributor system, in which a separate ignition coil generates the high voltage for each cylinder.

In addition to systems with electronic multipoint fuel injection, the following simpler and, at that time, cost-effective systems were also developed to allow Motronic to be used in mid-range and compact vehicles:
– KE-Motronic, based on KE-Jetronic continuous gasoline injection,
– Mono-Motronic, based on Mono-Jetronic intermittent single point injection.

State of the art
Now, only multipoint fuel-injection systems and fully electronic ignition are used for new vehicles. The following Motronic versions are used:
– ME-Motronic with electronic throttle control (ETC) for controlling injection, ignition, and fresh-air charge on intake-manifold injection systems (Figure 4).
– DI-Motronic (Direct Injection), with additional open- and closed-loop control functions for the high-pressure fuel circuit on gasoline direct-injection systems and to implement the various operating modes of this engine type (Figure 5 for homogeneous operation, Figure 6 for lean-burn concept).
– Bifuel-Motronic for controlling components for running the spark-ignition engine on natural gas or gasoline (see Engines fueled by natural gas).

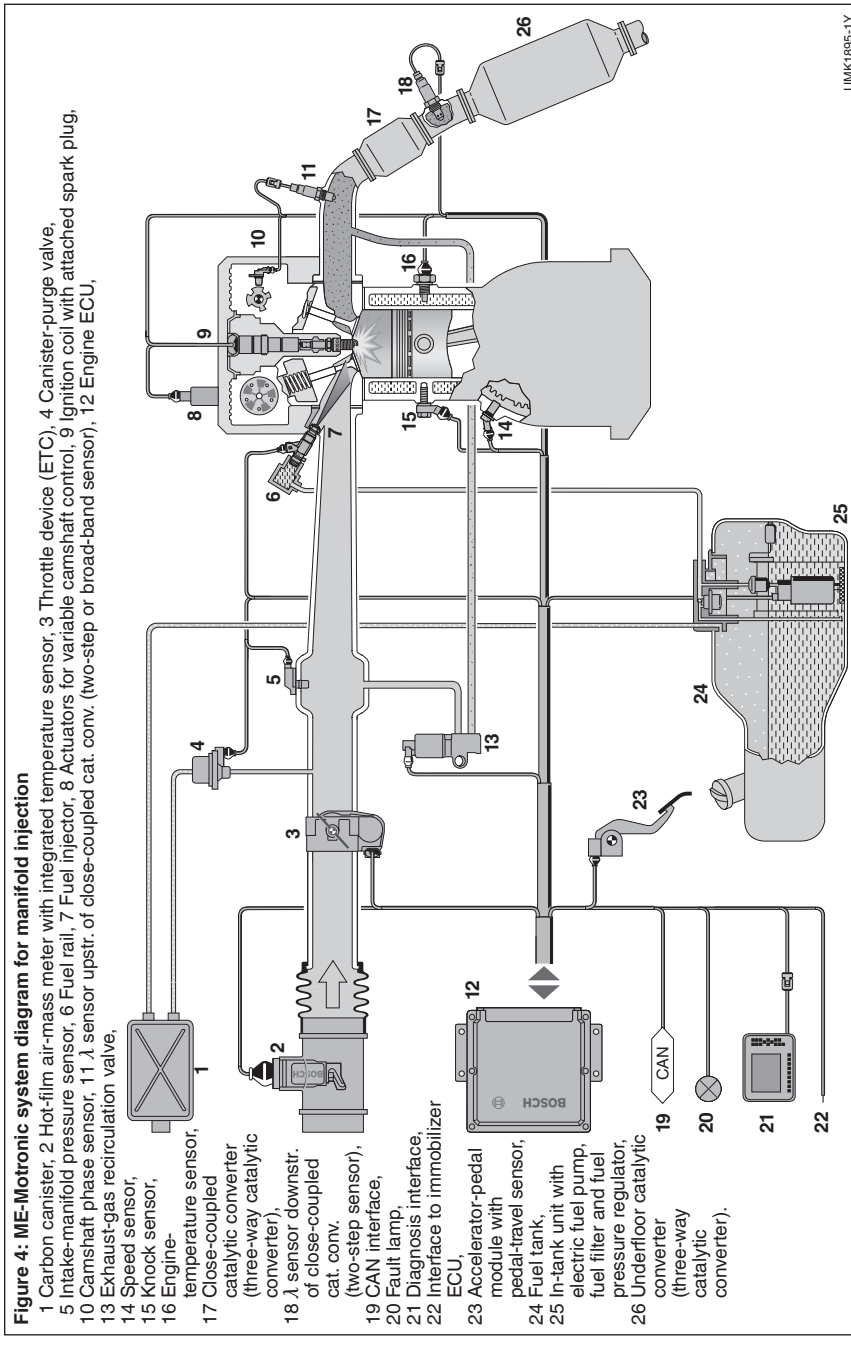


Figure 4: ME-Motronic system diagram for manifold injection
1 Carbon canister, 2 Hot-film air-mass meter with integrated temperature sensor, 3 Throttle device (ETC), 4 Canister-purge valve, 5 Intake-manifold pressure sensor, 6 Fuel rail, 7 Fuel injector, 8 Actuators for variable camshaft control, 9 Ignition coil with attached spark plug, 10 Camshaft phase sensor, 11 λ sensor upstr. of close-coupled cat. conv. (two-step or broad-band sensor), 12 Engine ECU, 13 Exhaust-gas recirculation valve, 14 Speed sensor, 15 Knock sensor, 16 Engine-temperature sensor, 17 Close-coupled catalytic converter (three-way catalytic converter), 18 λ sensor downstr. of close-coupled cat. conv. (two-step sensor), 19 CAN interface, 20 Fault lamp, 21 Diagnosis interface, 22 Interface to immobilizer ECU, 23 Accelerator-pedal module with pedal-travel sensor, 24 Fuel tank, 25 In-tank unit with electric fuel pump, fuel filter and fuel pressure regulator, 26 Underfloor catalytic converter (three-way catalytic converter).

Figure 5: DI-Motronic system diagram for gasoline direct injection (homogeneous operation) with exhaust-gas turbocharging

1 Carbon canister, 2 Canister-purge valve, 3 Non-return valves, 4 Blow-off valve, 5 Hot-film air-mass meter with integrated temperature sensor,
6 Charge-air pressure sensor (optionally in combination with charge-air temperature sensor), 7 Intercooler,
8 High-pressure pump with integrated fuel-supply control valve, 9 Actuators for variable camshaft control, 10 Ignition coil with attached spark plug,
11 Vacuum pump for wastegate activation, 12 Camshaft phase sensor, 13 Ambient-pressure sensor, 14 Throttle device (ETC),
15 Swirl control valve, 16 Fuel-pressure sensor, 17 Fuel rail (high-pressure rail), 18 High-pressure fuel injector, 19 Engine-temperature sensor,
20 Exhaust-gas temperature sensor, 21 Boost-pressure control valve, 22 Wastegate (bypass valve), 23 Exhaust-gas turbocharger,
24 Engine ECU,
25 Speed sensor,
26 Knock sensor,
27 λ sensor upstr. of close-coupled cat. conv. (two-step or broad-band sensor),
28 CAN interface,
29 Fault lamp,
30 Diagnosis interface,
31 Interface to immobilizer ECU,
32 Accelerator-pedal module with pedal-travel sensor,
33 Fuel tank,
34 In-tank unit with electric fuel pump, fuel filter and fuel pressure regulator,
35 Close-coupled catalytic converter (three-way catalytic converter),
36 λ sensor downstr. of close-coupled cat. conv. (two-point sensor),
37 Underfloor catalytic converter (three-way catalytic converter).

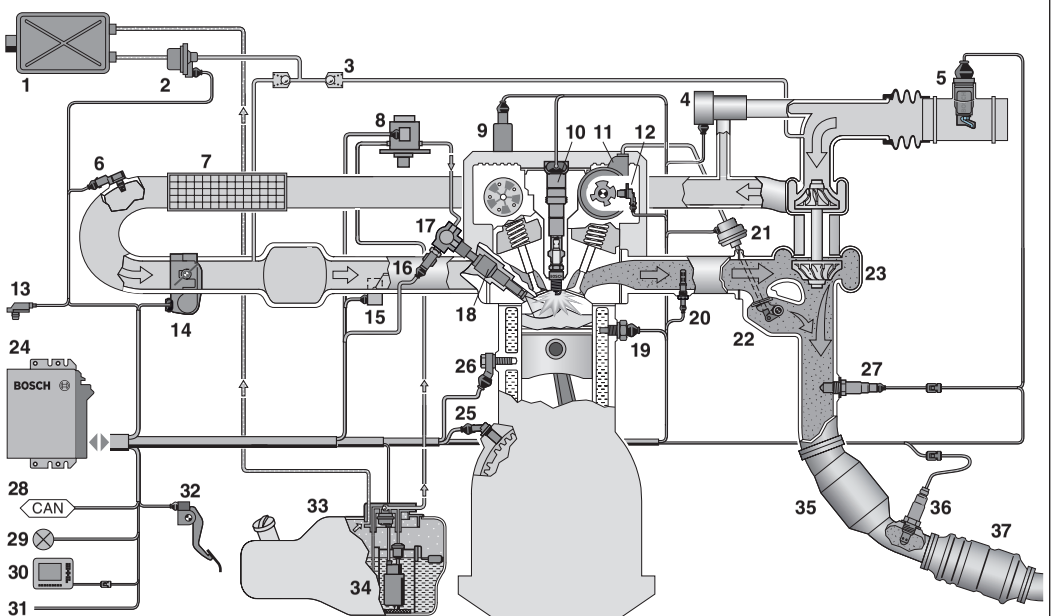
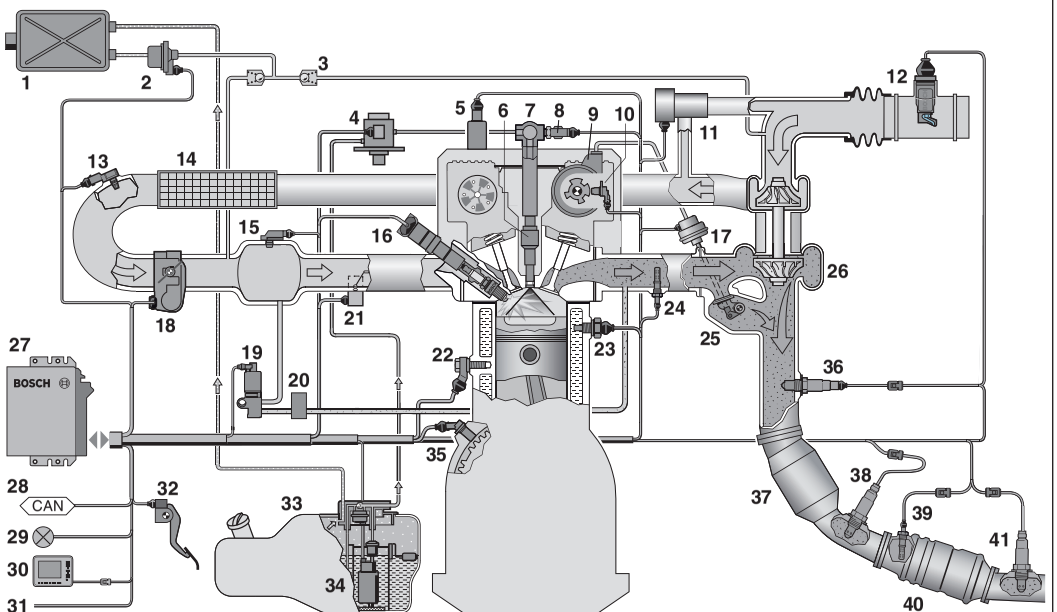


Figure 6: DI-Motronic system diagram for gasoline direct injection (lean-burn concept) with exhaust-gas turbocharging

1 Carbon canister, 2 Canister-purge valve, 3 Non-return valves, 4 High-pressure pump with integrated fuel-supply control valve, 5 Actuators for variable camshaft control, 6 High-pressure fuel injector (central arrangement for spray-guided combustion process), 7 Fuel rail (high-pressure rail), 8 Fuel-pressure sensor, 9 Vacuum pump for wastegate activation, 10 Camshaft phase sensor, 11 Blow-off valve, 12 Hot-film air-mass meter with integrated temperature sensor, 13 Charge-air pressure sensor (optionally in combination with charge-air temperature sensor), 14 Intercooler, 15 Intake-manifold pressure sensor, 16 Ignition coil with attached spark plug, 17 Boost-pressure control valve, 18 Throttle device (ETC), 19 Exhaust-gas recirculation valve, 20 Pressure sensor, 21 Swirl control valve, 22 Knock sensor, 23 Engine-temperature sensor, 24 Exhaust-gas temperature sensor, 25 Wastegate (bypass valve), 26 Exhaust-gas turbocharger, 27 Engine ECU, 28 CAN interface, 29 Fault lamp, 30 Diagnosis interface, 31 Interface to immobilizer ECU, 32 Accelerator-pedal module with pedal-travel sensor, 33 Fuel tank, 34 In-tank unit with electric fuel pump, fuel filter and fuel pressure regulator, 35 Speed sensor, 36 λ sensor upstr. of close-coupled cat. conv. (broad-band sensor), 37 Close-coupled catalytic converter (three-way catalytic converter or combination of three-way and NO_x storage catalytic converter), 38 λ sensor downstr. of close-coupled cat. conv. (two-step sensor), 39 Exhaust-gas temperature sensor, 40 Underfloor catalytic converter (NO_x storage catalytic converter), 41 λ sensor (two-step sensor) or NO_x sensor.



Motronic versions for motorcycles

The huge diversity of motorcycles available on the market internationally is reflected in the wide range of requirements imposed on engine management. Engine management has not by a long shot arrived on the scene in all segments and markets of the motorcycle sector. The Asian market in particular – India and China – is dominated by single-cylinder vehicles with low engine displacement under 150 cm^3 which are still entirely reliant on carburetor technology. Today, the persistent widespread use of air-cooled engines (see Engine cooling) is a particular characteristic of the motorcycle market.

Because of stricter, applicable legislation for example in Europe (see Emission-control and diagnosis legislation) motorcycles under 150 cm^3 already have electronic engine-management systems.

Motronic for multiple-cylinder engines
Adapted engine-management systems from the passenger-car sector are used for large-capacity multiple-cylinder engines in motorcycles.

Motronic for single-cylinder engines
For single-cylinder engines there are standalone solutions for engine management. A distinction is made between the following systems:
– MSE-M Motronic (SE, Small Engines) for controlling ignition and fuel injection on intake-manifold injection systems with mechanical throttle valves.
– MSE-E Motronic for controlling ignition, fuel injection and fresh-air charge on intake-manifold injection systems with electronic throttle valves (ETC).

Two particular features of these single-cylinder engines for motorcycles are the high rotational non-uniformities of the crankshaft (Figure 7) and the high intake manifold-pressure dynamics over a working cycle (see Cylinder charge). This behavior characteristic of single-cylinder engines can be used to implement the assignment of the engine position to the compression cycle or to the exhaust cycle without a camshaft speed sensor.

System simplifications in the small-engine segment play a key role in equipping the huge number of motorcycles in Asia with efficient engine management.

Requirements
A further requirement that sets apart engine management for motorcycles – both single- and multiple-cylinder motorcycles – from that for passenger cars is the higher rpm range to be covered between idle and maximum speeds. This imposes high demands on the computing capacity of Motronic.

The increasing requirements with regard to exhaust-gas quality and requirements from on-board diagnosis (see Emission-control and diagnosis legislation) will result in system-engineering changes in the future. Thus, a second, heated λ sensor downstream of the catalytic converter is used for example to diagnose catalytic-converter aging. The catalytic converter itself must be adapted in terms of its dimensions to the more stringent limits. The catalytic-converter heating strategy performed by the engine management makes an important contribution to heating the catalytic converter quickly.

Figure 7: Comparison of rotational non-uniformity for cars and motorcycles.
1 Three-cylinder engine for cars,
2 Single-cylinder engine for motorcycles.

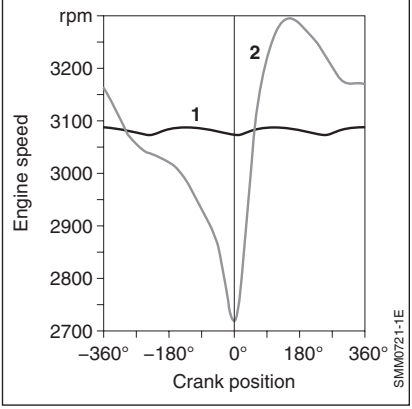


Figure 8: MSE-M Motronic with mechanical throttle valve for motorcycles
(System diagram with air-cooled engine).
1 Carbon canister, 2 Tank vent valve,
3 In-tank unit with electric fuel pump, fuel filter and fuel pressure regulator,
4 Fuel injector, 5 Ignition coil, 6 Spark plug, 7 rpm sensor,
8 Engine-temperature sensor, 9 λ sensor upstream of catalyst,
10 Intake-manifold pressure and temperature sensor, 11 Engine control unit, 12 CAN interface,
13 Fault lamp, 14 Diagnosis interface, 15 Mechanical throttle device,
16 Throttle grip with Bowden cable, 17 Throttle-valve sensor, 18 Controlled air bypass.

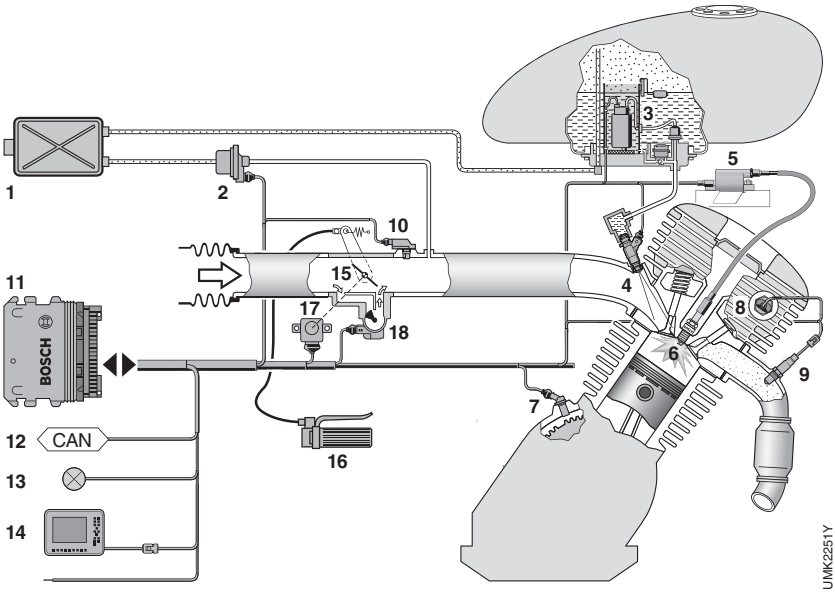
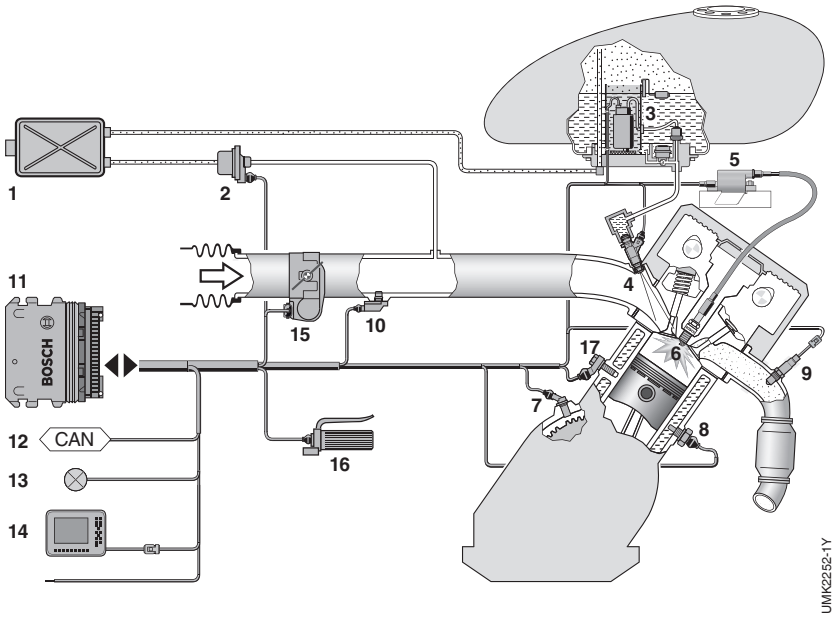


Figure 9: MSE-E Motronic with electronic throttle valve (ETC) for motorcycles
(System diagram with water-cooled engine).
1 Carbon canister, 2 Tank vent valve,
3 In-tank unit with electric fuel pump, fuel filter and fuel pressure regulator,
4 Fuel injector, 5 Ignition coil, 6 Spark plug, 7 rpm sensor,
8 Engine-temperature sensor, 9 λ sensor upstream of catalyst,
10 Intake-manifold pressure and temperature sensor, 11 Engine control unit, 12 CAN interface,
13 Fault lamp, 14 Diagnosis interface, 15 Electronic throttle device (ETC),
16 Throttle grip with grip travel sensor, 17 Knock sensor.



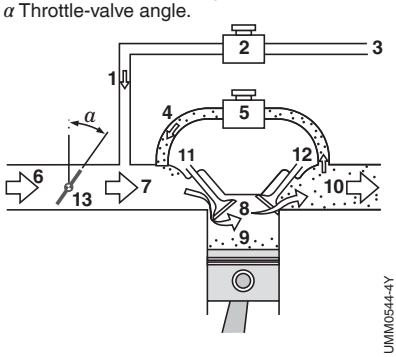
Cylinder charge

Component parts

The gas mixture located in the cylinder after the intake valves have closed is termed the cylinder charge. It consists of the supplied fresh A/F mixture and the residual exhaust gas. The major components that influence the cylinder charge are shown in Figure 1. The throttle valve is the most important component on modern systems.

Fresh A/F mixture
The component parts of the fresh A/F mixture drawn in are fresh air and – on systems with external mixture formation – the fuel suspended in it (Figure 1). Most of the fresh air flows through the throttle valve; additional fresh A/F mixture can be drawn in through the evaporative-emissions control system. The air mass present in the cylinder after the intake valves have closed is the decisive factor in the work performed above the piston during combustion, and ultimately in the torque delivered by the engine. Therefore, measures for increasing maximum torque and maximum engine power almost always require an increase in the maximum possible cylinder charge. The theoretical maximum charge is predetermined by the piston displacement and, in the case of turbocharged engines, by the achievable charge-air pressure, as well.

Figure 1: Cylinder charge in a spark-ignition engine
1 Air and fuel vapor,
2 Canister-purge valve (tank vent valve),
3 Connection to evaporative-emissions control system,
4 Exhaust gas,
5 Exhaust-gas recirculation valve (EGR valve),
6 Air-mass flow (ambient pressure),
7 Air-mass flow (intake-manifold pressure),
8 Fresh A/F mixture charge (combustion-chamber pressure),
9 Residual exhaust-gas charge (combustion-chamber pressure),
10 Exhaust gas (exhaust-gas backpressure),
11 Intake valve,
12 Exhaust valve,
13 Throttle valve (butterfly valve).
α Throttle-valve angle.



Residual exhaust gas
The residual exhaust gas in the charge is formed by:
– the mass of exhaust gas that remains in the cylinder or is stored briefly during the charge cycle in the intake or exhaust port and then drawn back again, and is not ejected on completion of the charge cycle via the exhaust valve, and
– in the case of systems with exhaust-gas recirculation by the mass of the returned exhaust gas.

The amount of residual exhaust gas is determined by the charge cycle. It does not contribute directly to the combustion process, but does influence ignition and the combustion process as a whole. At wide-open throttle, the amount of residual exhaust gas generally needs to be as small as possible in order to maximize the fresh-air mass and also the power output of the engine.

At part load, however, the amount of residual exhaust gas is desirable in order to reduce fuel consumption. This is achieved by a more favorable cycle, resulting from a change in mixture composition, and also by a reduction in pump losses during the charge cycle, as a higher intake-manifold pressure is required for the same air charge. A specifically introduced amount of residual exhaust gas can likewise reduce the emission of nitrogen oxides (NO_x) and unburned hydrocarbons (HC).

Control of cylinder charge

On spark-ignition (SI) engines with external mixture formation (intake-manifold injection) and on systems with internal mixture formation (gasoline direct injection) and homogeneous cylinder charge, the torque delivered by the engine is determined by the air charge. In contrast to this, in the case of internal mixture formation with an air surplus, the engine torque can also be controlled directly by varying the injected fuel mass (stratified-charge operation).

Throttle valve

The critical adjustment mechanism for controlling air-mass flow supplied to the cylinders is the throttle valve. When the throttle valve is not fully open, the air flow drawn in by the engine is throttled, thus reducing maximum engine torque. This throttling effect is dependent on the position, and thus the opening cross-section of the throttle valve, as well as on engine speed (Figure 2). Maximum torque is developed with the throttle valve wide open.

By the end of the 1990s the mechanical throttle valve was state of the art. The driver directly controls the opening of the throttle valve via a Bowden cable by pressing the accelerator pedal.

On systems with electronic throttle control (ETC), the required air charge is calculated from the desired engine torque, which is obtained from the accelerator-pedal position, and the throttle valve is activated accordingly.

Electronic throttle control (ETC) components

The electronic throttle control system (Figure 3) consists of the accelerator-pedal module, the engine control unit, and the throttle device. The throttle device mainly consists of the throttle valve, the electric throttle-valve drive element, and the throttle-valve position sensor. The drive element is a DC motor, which acts on the throttle-valve shaft via a gear unit. The throttle-valve position sensor, designed as a redundant unit, detects the position of the throttle valve.

The driver command is detected by the redundant sensor system in the accelerator-pedal module, and the signal is sent to the engine control unit. This calculates the required cylinder charge, based on the current engine operating point, and regulates the opening angle of the throttle

Figure 2: Throttle map of a spark-ignition engine

- - - Intermediate position of the throttle valve.

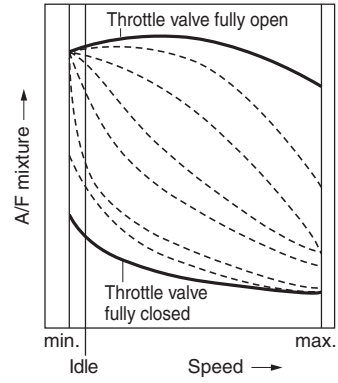
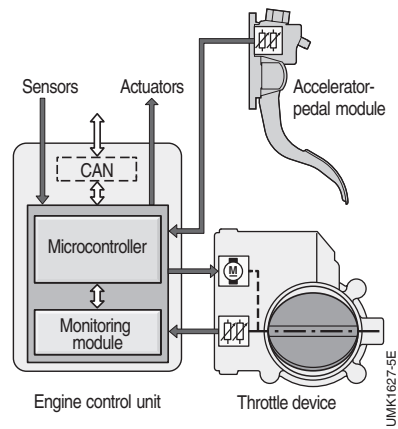


Figure 3: ETC system



valve by means of the throttle-valve drive element and the throttle-valve position sensor.

The redundancies incorporated in the accelerator-pedal module and the throttle device are part of the ETC monitoring concept for avoiding malfunctions.

Charge cycle

The charge cycle of fresh A/F mixture and residual exhaust gas is controlled by the opening and closing of the intake and exhaust valves. The opening and closing times of the valves (timing) and the valve lift curve are critical factors.

Valve timing

Valve overlap, i.e. the overlap of the opened times of the intake and exhaust valves, has a decisive influence on the mass of residual outlet gas remaining in the cylinder. The quantities of fresh A/F mixture and residual exhaust gas in the cylinder can be controlled by changing the valve-lift time curves.

In the overlap phase exhaust gas flows through the intake valve into the intake system due to the pressure differential between the exhaust system and the intake manifold. In the induction stroke it then returns to the combustion chamber (internal exhaust-gas recirculation). The longer the valve overlap lasts (early opening of the intake valve), the more mass of exhaust gas remains in the cylinder. This results in shortened combustion temperatures and thus a reduction in NO_x emissions.

However, since recirculated exhaust gas displaces fresh A/F mixture, opening the intake valve early also leads to a reduction in the maximum torque. In addition, excessively high exhaust-gas recirculation, particularly at idle, can lead to combustion misses which, in turn, cause an increase in HC emissions. With variable valve timing, varying the timing dependent on the operating point can optimize emissions.

By varying the valve timing accordingly (e.g. continuous adjustment of the phase position and valve lift), it is possible to control the air-mass flow, and thus engine performance, without having to use the throttle valve. The amount of residual exhaust gas can also be adjusted by means of valve timing.

On today's systems, the valves are actuated mechanically via the camshaft. Additional measures can be used to vary this activation to a certain degree (e.g. camshaft adjustment or camshaft shifting). However, these mechanical systems cannot dispense with the throttle valve completely.

Scavenging

In modern turbocharged engines valve-gear variabilities are additionally used to increase significantly the torque in the low-rpm range. This timing strategy is called scavenging. For this purpose the phase position of the intake valves and if necessary the exhaust valves are adjusted in the full-load range in such a way as to produce a valve overlap. During the valve overlap a partial flow of air passes directly again through the exhaust valve to the exhaust side, improving the scavenging of residual gas and with it enhancing anti-knock performance. At the same time the additional scavenging air improves process control at the turbocharger so that higher charge-air pressures can be achieved.

Exhaust-gas recirculation (EGR)

The mass of residual exhaust gas in the cylinder can also be increased by "external exhaust-gas recirculation" (EGR). In this case, an exhaust-gas recirculation valve connects the intake manifold and exhaust manifold (Figure 1). When the valve is open and with a pressure differential between the exhaust-gas system and the intake manifold, the engine draws in a combination of fresh A/F mixture and exhaust gas. The engine ECU calculates the level of exhaust-gas recirculation for a specific operating state, and activates the exhaust-gas recirculation valve accordingly.

Fuel reduction

Exhaust-gas recirculation increases intake-manifold pressure. The higher intake-manifold pressure leads to a reduction in the charge cycle work and this lowers fuel consumption.

NO_x reduction
Exhaust-gas recirculation is used on engines with gasoline direct injection which operate in lean-burn mode (stratified-charge operation) to reduce NO_x emissions. Exhaust-gas recirculation is the main way to minimize untreated NO_x emissions, and thus prolong the duration of lean-burn operation, which is limited by the NO_x storage catalytic converter.

Exhaust gas can be conducted back to the combustion chamber to reduce peak combustion temperatures. The reduction in temperature is based on the fact that the returned exhaust gas does not participate in the combustion and thus delivers no combustion energy. It is also an additional thermal mass, which means that the combustion energy distributes itself to a higher overall mass.

Higher combustion temperatures induce an over-proportional increase in NO_x formation, and as exhaust-gas recirculation (EGR) reduces combustion temperatures, it represents a particularly effective means of controlling NO_x emissions.

Supercharging
The torque obtainable is proportional to the charge of fresh A/F mixture. This enables an increase in the maximum torque in that the air in the cylinder is compressed by means of dynamic charging, mechanical charging or exhaust turbocharging (see Supercharging).

Recording of air charge

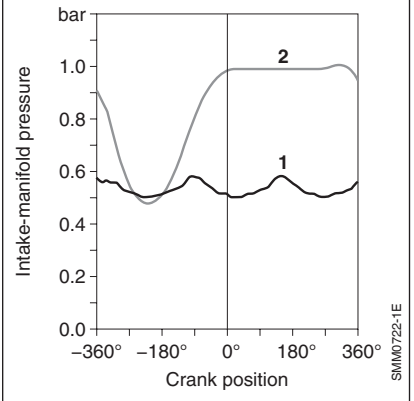
The cylinder charge crucially influences the power output, fuel consumption, and emissions. The composition of the air/fuel mixture plays a decisive role here (see Mixture formation, see Catalytic exhaust-gas treatment). To be able to adjust this correctly, it is necessary among others things to record the cylinder charge in the engine-management system with great precision. The throttle valve, as the central adjustment mechanism, determines the pressure in the intake manifold and uses this to influence the air-mass flow. Therefore the intake-manifold pressure is one of the central input variables which the engine management draws on to calculate the cylinder charge. The fresh-air mass however still depends on many additional factors, such as the ambient pressure, the throttle-valve leakage, the position of the gas-exchange valves, the operating status of the exhaust-gas turbocharger, the position of the swirl control valves, and the exhaust-gas recirculation. To take these influences into account, the engine-management system has in part further sensor technology, such as a hot-film air-mass meter to determine the cylinder charge.

Particularities for motorcycles
Motorcycles generally have less variability in the air system. Turbocharging, exhaust-gas recirculation and variable valve timing have barely featured up to now. Variable intake-manifold geometries have hitherto been the preserve of vehicles of higher performance classes. Particularly in small single-cylinder engines, therefore, the air charge is recorded purely by way of the throttle-valve position and intake-manifold pressure.

Compared with typical passenger-car applications, there are systematic differences in the intake manifold-pressure behavior. These are primarily attributable to different, geometric conditions. While in passenger-car applications a common throttle valve is for the most part used for a number of cylinders, motorcycles frequently rely on individual throttle valves (see Motorcycle intake manifolds). These are installed close to the cylinder head to guarantee good response. The absence of a classic manifold means that the volumes between the throttle valve and the intake valves are comparatively small. This results in small ratios of intake-manifold volume to displacement of the individual cylinder. This factor influences the intake manifold-pressure behavior (see Figure 4). While for a car a virtually constant vacuum is obtained in the intake manifold with a comparable load point and comparable engine speed, for a motorcycle the vacuum is achieved only briefly during the intake phase. As soon as the intake valves are closed, fresh air flows through the throttle valve and the intake-manifold pressure increases to ambient pressure. This behavior must be taken into account in the recording of charge by the engine management since the mean intake-manifold vacuum can only be used to a limited extent to determine the fresh-air mass.

Figure 4: Pressure characteristic in the intake manifold of car and motorcycle at part load

1 Three-cylinder car engine,
2 Single-cylinder motorcycle engine.

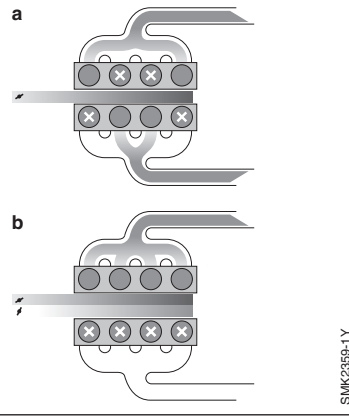


Cylinder shutoff

On engines with cylinder shutoff individual cylinders can be shut off at part load. The cylinders that continue to fire then generate the entire demanded torque. The air charge required for this is roughly doubled when every second cylinder is shut off referred to the corresponding full-engine operating point. The lower gas-exchange losses as a result of the de-throttling of the fired cylinders give rise to reduced fuel consumption. The gas-exchange losses can be further reduced by shutting off the gas-exchange valves of the non-fired cylinders. This shutoff necessitates accordingly designed valve gear (see Valve gear).

The most important motivation behind an engine design with cylinder shutoff is the reduction of fuel consumption. Since diesel engines are as a rule operated under de-throttled conditions to a large extent, the advantage is confined to the shutoff of the charge cycle. The reduced consumption that can be achieved by cylinder shutoff is dependent, aside from the operating-point profile, on the individual engine and drivetrain design. In gasoline engines with charge-cycle shutoff, it is realistic to achieve improvements in the higher single-digit percentage range in the conventional driving cycles and in the double-digit percentage range at individual operating points at low load.

Figure 5: Configuration of cylinder shutoff
a) Half-engine operation,
b) Cylinder-bank shutoff.



Configuration of cylinder shutoff
A cylinder shutoff can be implemented both in an engine with multiple cylinders and in an engine with only two cylinders. Different cylinder-shutoff configurations are possible here. The possible configurations are derived from the engine's design features. The most important criteria are the firing order and the respective layout of the intake and exhaust tracts. In most series applications every second cylinder in firing order and consequently half the cylinders is shut off. In half-engine operation (Figure 5a) shuttable-off and permanently fired cylinders are operated in the same bank. Here it is necessary to shut off the charge cycle, since λ control and the conversion capability of the catalytic converter would not be guaranteed with the oxygen that otherwise enters the exhaust tract.

When all the cylinders in an exhaust bank are shut off (cylinder-bank shutoff, Figure 5b), charge-cycle shutoff can also be dispensed with. This reduces the design complexity and the costs, but does not exploit the full CO_2 savings potential of cylinder shutoff.

Adaptation of vehicle components
With regard to NVH aspects (Noise Vibration Harshness), it may, aside from the valve gear, also be necessary to adapt the design of the dual-mass flywheel, the engine bearings, and the exhaust-gas system. An intake-manifold pressure sensor may also be needed to diagnose the valve actuators.

Engine management
Engine management is particularly important when it comes to operating an engine with cylinder shutoff. Partial-engine operation necessitates the adaptation of numerous software functions (Figure 6). Furthermore, additional functions for requirement and changeover coordination and diagnosis of the valve actuators for zero stroke are required.

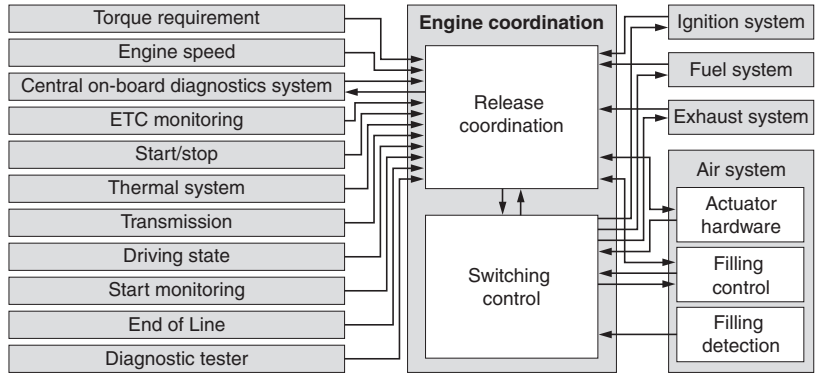
Control of the changeover between full- and partial-engine operation is particularly complex because all the changeovers must take place imperceptibly to the driver and the response must not be delayed compared with full-engine operation. As well as charge adaption, exactly timed fading-out and -in of fuel injection for the shuttable-off cylinders and interventions via the moment of ignition are required for this purpose. Therefore, as well as engine coordination, torque structure, air, fuel, ignition and exhaust systems, and ETC monitoring are also adapted.

Cylinder-shutoff requirement
Operating-mode control by engine management constantly monitors numerous conditions for cylinder shutoff. Among the most important input variables is the setpoint torque to be developed by the engine, since the maximum settable torque is limited with cylinder shutoff. Another important input variable is the engine speed, since reliable switching of the gas-exchange valves is not possible above a limit speed. Further input variables are the engine temperature and the driving speed. Conditions are assigned to each input variable. On this basis either partial- or full-engine operation is requested.

Changeover between partial- and full-engine operation
The changeover to partial-engine operation begins with the increase of the air charge so that the still-fired cylinder can generate the entire demanded engine torque. In an engine with one intake manifold the increase is effected jointly for all the cylinders. In an engine with separate intake manifolds the air change can be changed for the still-fired cylinder bank regardless of the bank to be shut off. To deliver the required setpoint torque, the moment of ignition must be retarded accordingly. As soon as a sufficient torque reserve has been created by way of the air-charge buildup, the gas-exchange valves can be deactivated for the cylinders to be shut off and fuel injection faded out. The moment of ignition is then advanced again.

The changeover to full-engine operation begins with the activation of charge cycle and fuel injection and ignition-angle adjustment. The moment of ignition is advanced in parallel with the charge adaptation.

Figure 6: Overview of the coordination of engine management for cylinder shutoff



Fuel supply

Fuel supply and delivery with intake-manifold injection

Different system configurations operating with intake-manifold injection typically at fuel pressures of approx. 300 to 400 kPa (3 to 4 bar) are used to supply fuel.

System with fuel return
The electric fuel pump delivers the fuel and generates the injection pressure (Figure 1). The fuel is drawn from the fuel tank and passes through the fuel filter into the high-pressure line, from where it flows to the engine-mounted fuel rail (fuel distributor). The rail supplies the fuel to the fuel injectors.

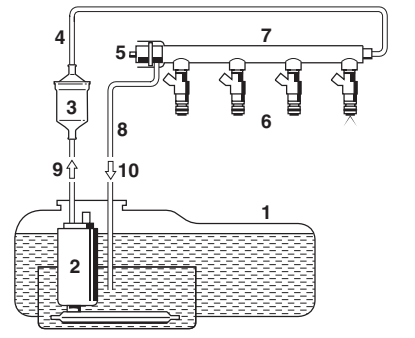
The mechanical fuel-pressure regulator mounted on the rail is a differential-pressure regulator with pneumatic connection to the intake manifold. It keeps the difference between fuel pressure in the rail and the intake-manifold pressure constant. The amount of fuel metered by the fuel injectors depends on this pressure differential because fuel is injected against the pressure in the intake manifold. Because the pressure regulator regulates to a constant pressure differential irrespective of the absolute intake-manifold pressure – i.e., irrespective of the engine load – the amount of fuel injected is only dependent on the duration of the fuel-injector activation time (injection time).

The fuel not needed by the engine flows through the rail via the return line connected to the fuel-pressure regulator back to the fuel tank. This excess fuel heated in the engine compartment causes the fuel temperature in the fuel tank to rise. Fuel vapor is generated as a function of fuel temperature. Ensuring adherence to environmental-protection regulations, the vapors are routed through a tank-ventilation system for intermediate storage in a carbon canister until they can be returned through the intake manifold for combustion in the engine (see Evaporative-emissions control system).

Returnless system
In a returnless fuel-supply system the fuel-pressure regulator is located in the fuel-supply module installed in the fuel tank, which means that the return line from the engine to the fuel tank is no longer required (Figure 2). Since the fuel-pressure regulator on account of its installation location has no reference to the intake-manifold pressure, the relative injection pressure compared with the intake-manifold pressure is not dependent on the engine load. This is taken into account in the calculation of the injection time in the engine control unit.

Only the amount of fuel which is to be injected is delivered to the rail. The excess flow volume delivered by the elec-

Figure 1: Fuel supply and delivery with intake-manifold injection (system with fuel return)
1 Fuel tank,
2 Electric fuel pump,
3 Fuel filter,
4 Pressure line,
5 Fuel-pressure regulator,
6 Fuel injectors,
7 Rail (fuel flowing through),
8 Return line,
9, 10 Flow direction of fuel.



tric fuel pump returns directly to the fuel tank without taking the circuitous route through the engine compartment. In this way, fuel heating in the fuel tank and thus also evaporative emissions are significantly lower than in systems with fuel return. Because of these advantages, it is mainly only returnless systems which are now used.

Demand-regulated returnless system

In a demand-regulated system, the fuel-supply pump delivers only that amount of fuel that is currently used by the engine and that is required to set up the desired pressure. Pressure control is effected by means of a closed control loop in the engine control unit, whereby the current fuel pressure is recorded by a pressure sensor (Figure 2). This dispenses with the mechanical fuel-pressure regulator. To vary the delivery quantity the triggering power of the electric fuel pump is altered by means of a clock module that is triggered by the engine control unit; the triggering is performed with pulse-width modulation (PWM).

The system is equipped with a pressure-relief valve to prevent the buildup of excessive pressure even during overrun fuel cutoff or after the engine has been switched off.

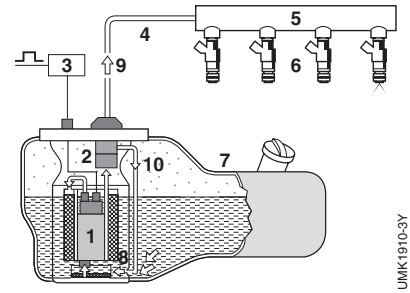
As a result of demand control, no excess fuel is compressed and thus the capacity of the electric fuel pump is minimized. This results in reduced fuel consumption compared with system with non-controlled electric fuel pumps. The in-tank fuel temperature can also be further reduced.

Further advantages of a demand-regulated system are derived from the variably adjustable fuel pressure. On the one hand, the pressure can be increased during hot starting to prevent the formation of vapor bubbles. On the other hand, it is possible above all in turbocharging applications to meter better the very large injected quantities required at high loads and the very small quantities at idle (injected-fuel-quantity spread) by effecting a pressure increase at full load and a pressure decrease at very low loads. This resolves the problem of metering small quantities in the fuel-injector configuration.

Furthermore, the measured fuel pressure provides for improved diagnostic options for the fuel system compared with previous systems. More precise fuel metering is achieved by taking into account the current fuel pressure when calculating the injection time.

Figure 2: Fuel supply and delivery with intake-manifold injection (demand-regulated returnless system)

1 Electric fuel pump with fuel filter (fuel filter may also be mounted outside of the fuel tank),
2 Pressure-relief valve and pressure sensor (pressure sensor alternatively in the rail),
3 Clock module for controlling the electric fuel pump,
4 Pressure line, 5 Rail (returnless),
6 Fuel injectors, 7 Fuel tank,
8 Suction-jet pump,
9, 10 Flow direction of fuel.



Fuel supply and delivery with gasoline direct injection

Compared with injecting fuel into the intake manifold, there is only a limited time window available for injecting fuel directly into the combustion chamber. Mixture formation is therefore more important and fuel must be injected in the case of direct injection at a pressure approximately 50 times greater than is the case with intake-manifold injection.

The fuel system is divided into a low-pressure circuit and a high-pressure circuit.

Low-pressure system

With gasoline direct injection the low-pressure system serves to supply the high-pressure system, involving the use of the same fuel systems and components as intake-manifold injection. Because of the high temperatures at the high-pressure pump in hot-start and hot-running conditions, a higher predelivery pressure (admission pressure) is required to prevent the formation of vapor bubbles. It is therefore an advantage to use systems with variable low pressure. Demand-regulated low-pressure systems are therefore particularly suitable here, because they allow the optimum admission pressure to be set for every engine operating state; this admission pressure usually varies in a range of 300 to 600 kPa (3 to 6 bar) relative to the ambient pressure.

Demand-controlled systems are increasingly being used as a further variant. In contrast to demand-regulated systems, in which the pressure is adjusted via a closed control loop by the pressure sensor, the pressure sensor is omitted here and the system costs are thus reduced. The electric fuel pump is activated 100 % by pilot control as a function of the engine operating point.

High-pressure system

High-pressure fuel circuits use predominantly demand-regulated high-pressure pumps as well as continuous-delivery high-pressure pumps. However, continuous-delivery high-pressure pumps can still be found in first-generation gasoline direct-injection systems.

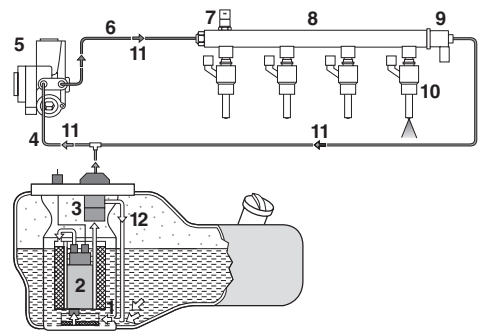
The high-pressure system comprises the fuel rail (high-pressure rail) with high-pressure injectors and a high-pressure sensor (Figures 3 and 4). The continuous-delivery system also requires a separate pressure-control valve.

Depending on the engine's operating point, the pressure varies typically in a continuous-supply system in a range of 5 to max. 11 MPa (50 to 110 bar), and in a demand-regulated system up to max. 35 MPa (350 bar). The high-pressure sensor provides the information on the currently prevailing pressure.

The signal from the high-pressure sensor is also used for injection calculation and for fuel-system diagnostics.

Figure 3: Fuel supply and delivery with gasoline direct injection (continuous-delivery system)

1 Suction-jet pump,
2 Electric fuel pump with fuel filter,
3 Pressure regulator,
4 Low-pressure line,
5 Continuous-delivery high-pressure pump,
6 High-pressure line,
7 High-pressure sensor,
8 Rail,
9 Pressure-control valve,
10 High-pressure injectors,
11, 12 Flow direction of fuel.



Continuous-delivery system
The high-pressure pump driven by the engine camshaft – normally a three-barrel radial-piston pump (see High-pressure pumps for gasoline direct injection) – forces fuel into the rail against the system pressure (Figure 3). The pump delivery quantity cannot be adjusted. The excess fuel not required for fuel injection and to maintain the pressure is depressurized by the pressure-control valve and returned to the low-pressure circuit. For this purpose, the pressure-control valve is actuated by the engine control unit in such a way as to obtain the injection pressure required at a given operating point. The pressure-control valve also serves as a mechanical pressure-relief valve.

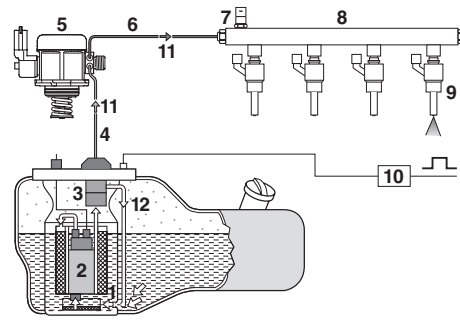
In continuous-delivery systems, most of the operating points cause significantly more fuel to be compressed to high system pressure than is needed by the engine. This results in unnecessary energy consumption and thus higher fuel consumption than demand-regulated systems. The excess fuel released via the pressure-control valve also contributes to increasing the temperature in the fuel system. For these reasons, only demand-regulated high-pressure systems are used in modern direct-injection engines.

Demand-regulated system
In a demand-regulated system (Figure 4), the high-pressure pump – usually a single-barrel radial-piston pump (see High-pressure pumps for gasoline direct injection) – delivers to the fuel rail only the precise amount of fuel which is actually needed for injection and to set the desired pressure. The pump is usually driven by the engine camshaft, single-barrel pumps being driven via special cams which drive the pump plunger. The delivery quantity is varied by a fuel-supply control valve which is integrated in the high-pressure pump. The engine control unit actuates this valve with each pump lift with such precision as to provide the required delivery quantity in order to set the system pressure needed in the rail for the given operating point (see Fuel-supply control valve).

For safety reasons the high-pressure circuit incorporates a mechanical pressure-relief valve, which is usually integrated directly in the high-pressure pump. Should the pressure exceed the permissible level, fuel is returned via the pressure-relief valve to the low-pressure circuit.

Figure 4: Fuel supply and delivery with gasoline direct injection (demand-regulated system)

1 Suction-jet pump,
2 Electric fuel pump with fuel filter,
3 Pressure-relief valve and pressure sensor (pressure sensor alternatively in the low-pressure line),
4 Low-pressure line,
5 Demand-regulated high-pressure pump with integrated fuel-supply control valve and pressure-relief valve,
6 High-pressure line,
7 High-pressure sensor,
8 Rail,
9 High-pressure injectors,
10 Clock module for controlling the electric fuel pump,
11, 12 Flow direction of fuel.



Evaporative-emissions control system

Function

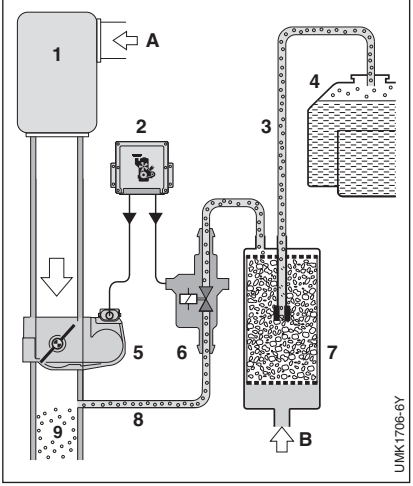
An evaporative-emissions control system (tank ventilation) is required for vehicles equipped with a gasoline engine. Its purpose is to intercept and collect fuel degassing from the fuel tank and to comply with the legal stipulations governing emission limits for evaporative loss. In addition, the line from the fuel tank via the carbon canister to atmosphere provides for pressure equalization of the fuel tank which is required on account of temperature-induced expansion of the fuel or the tank system itself or while the vehicle is being refueled. Degassing increases when the fuel in the fuel tank is heated. This may be caused by the higher ambient temperatures, the power loss of the fuel pump integrated in the fuel tank or – depending on the fuel supply system – the return of fuel heated in the engine and no longer required for the combustion process. Degassing also occurs when the ambient pressure drops, for example due to weather influences or when driving uphill.

Design and operating principle

The evaporative-emissions control system consists of a carbon canister, into which the vent line from the fuel tank projects, as well as a canister-purge valve that is connected both to the carbon canister as well as to the intake manifold (Figure 5). The activated carbon absorbs the fuel vapor. Due to the vacuum prevailing in the intake manifold, fresh air is drawn through the activated carbon when the canister-purge valve frees up the line between the carbon canister and the intake manifold when driving. The fresh air picks up the absorbed fuel and feeds it to the combustion process. This is know as purging the carbon canister.

The engine control unit controls the purge-gas volume depending on the engine operating point. In order to ensure that the carbon canister is always able to absorb fuel vapor, the activated carbon must be regenerated at regular intervals. Since current gasoline engines are operated to increase efficiency in such a way that a high intake-manifold vacuum is avoided (e.g., with downsizing, exhaust-gas turbocharging, lean-burn concepts), adequate regeneration is often not feasible. In this case an additional regenerating line is used which utilizes the pressure drop after the air filter for additional vacuum generation (Figure 6). In turbocharged engines a venturi tube can also be used for this purpose. This utilizes a small part of the charge-air pressure to generate a suction flow to the carbon canister using the loss of pressure at its constriction. A non-return valve is integrated in each ventilation path so that there are no circular flows.

Figure 5: Evaporative-emissions control system
1 Air filter,
2 Engine control unit,
3 Fuel-tank vent line,
4 Fuel tank,
5 Throttle device,
6 Canister-purge valve,
7 Carbon canister,
8 Line to intake manifold,
9 Intake manifold.
A Intake air,
B Fresh air.



Pressure tank
With hybrid vehicles, which can be run on purely electric power for longer distances, there is in some cases no further opportunity for regeneration. In this case, the usual fuel tank made of plastic is replaced by a pressure-resistant steel tank which is connected via a further shutoff valve mounted on the tank to the carbon canister (Figure 6). In the event of increased degassing the pressure in such a tank rises, which counteracts further degassing of the fuel. If the pressure in the tank reaches critical values, the tank pressure is regulated by the engine ECU, whereby the additional shutoff valve opens and the fuel vapors are channeled into the carbon canister. The tank pressure must also be reduced to refuel the vehicle. For regeneration, operation of the internal-combustion engine is then explicitly requested by the engine ECU.

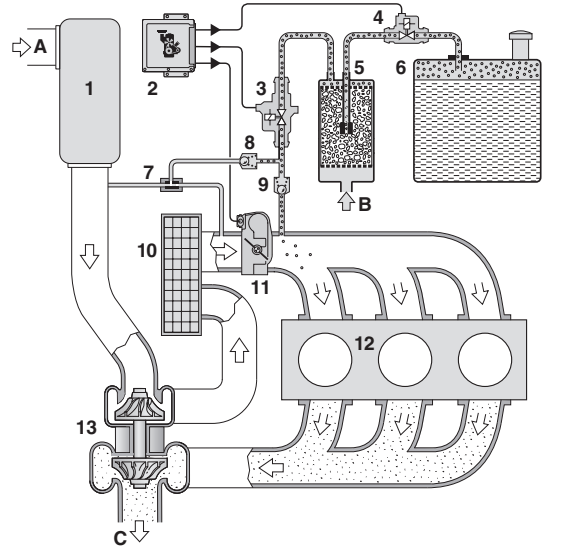
Gasoline filter

Function
The function of the fuel filter is to filter the fuel that flows into the fuel system. Contaminants in the fuel must be filtered out in order to protect the system and in particular the fuel injectors.

Design
Fuel filters for gasoline engines (gasoline filters) are located on the pressure side after the fuel-supply pump. In-tank filters are the preferred choice in newer vehicles, i.e., the filter is integrated in the fuel-supply module in the fuel tank. In this case, it must always be designed as a lifetime filter, which does not need to be changed over the full service life of the vehicle. Furthermore, in-line filters, which are installed in the fuel line, continue to be used. These can be designed as replacement parts or lifetime parts.

The filter housing is manufactured from steel, aluminum or plastic. It is connected to the fuel feed line by a thread, tube or quick-action connection. The housing contains the filter element, which filters

Figure 6: Evaporative-emissions control system with additional regenerating line
The figure shows an application for hybrid vehicles with a steel tank and an additional shutoff valve.
1 Air filter,
2 Engine control unit,
3 Canister-purge valve,
4 Shutoff valve,
5 Carbon canister,
6 Fuel tank (pressure-resistant steel tank),
7 Venturi tube,
8 Non-return valve,
9 Non-return valve,
10 Intercooler,
11 Throttle device,
12 Engine,
13 Exhaust-gas turbocharger.
A Intake air,
B Fresh air,
C Exhaust gas.



the dirt particles out of the fuel (Figure 7). The filter element is integrated in the fuel circuit in such a way that fuel passes through the entire surface of the filter medium as much as possible at the same flow velocity.

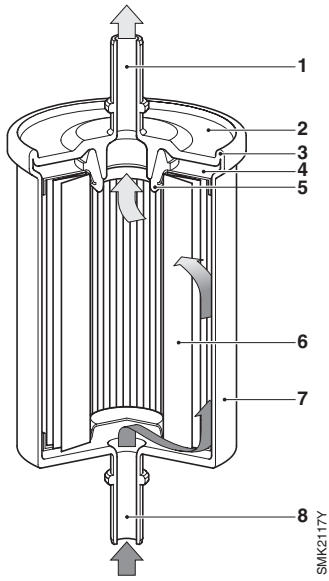
Filter medium

Special resin-impregnated cellulose-fiber papers which are also bonded for heavier-duty applications to a synthetic-fiber (meltblown) layer are used as the filter medium. This bond must ensure high mechanical, thermal and chemical stability. The paper porosity and the pore distribution of the filter paper determine the filtration efficiency and throughflow resistance of the filter.

Fuel filters for gasoline engines are either spiral vee-form or radial vee-form in design. In a spiral vee-form filter, an embossed filter paper is wrapped round a support tube. The fuel flows through the filter in the longitudinal direction.

Figure 7: In-line gasoline filter

1 Fuel outlet, 2 Filter cover,
3 Laser-welded edge, 4 Support ring,
5 Seal, 6 Radial vee-form filter medium,
7 Pressure-stable filter housing,
8 Fuel inlet.



In a radial vee-form filter, the filter paper is folded and inserted into the housing in the shape of a star (Figure 7). End rings made of plastic, resin or metal and, if necessary, an inner protective jacket provide stability. The fuel flows through the filter from the outside inwards, during which the dirt particles are separated from the filter medium.

Requirements

The required filter fineness is dependent on the fuel-injection system. For systems with intake-manifold injection, the filter element has a mean pore width of approx. 10 µm. Finer filtering is required for gasoline direct injection. The mean pore width here is in the range of 5 µm. Particles larger than 5 µm must be separated off by up to 85 %. In addition, a filter for gasoline direct injection, when new, must satisfy the following residual-dirt requirement: Metal, mineral and plastic particles and glass fibers with diameters of more than 400 µm must not be flushed by the fuel out of the filter.

Filter efficiency depends on the throughflow direction. When replacing in-line filters, it is imperative that the flow direction specified by the arrow be observed.

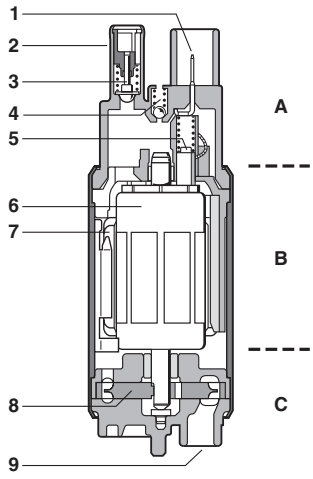
The interval for changing conventional in-line filters is, depending on filter volume and fuel contamination, normally between 30,000 km and 120,000 km. In-tank filters can generally be used for at least 250,000 km before they have to be replaced. This corresponds to today's design lifetime for a gasoline engine. There are in-tank and in-line filters available for use with gasoline direct-injection systems which feature service lives in excess of 250,000 km. If the in-tank filter has to be exchanged, this is only possible by completely changing the fuel-supply module.

Electric fuel pump

The electric fuel pump must deliver sufficient quantities of fuel to the engine and maintain the pressure required for efficient fuel injection in all operating states. Essential requirements include:
– Delivery quantity between 40 and 300 *l*/h at nominal voltage
– Pressure in the fuel system between 300 and 650 kPa (3.0 to 6.5 bar)
– Buildup of system pressure from 50 to 60 % of nominal voltage; the decisive factor here is operation during cold starting.

Apart from this, the electric fuel pump is increasingly being used as the pre-supply pump for the modern direct-injection systems used on diesel and gasoline engines. In the case of gasoline direct injection, sometimes pressures of up to 650 kPa must be provided during hot-delivery operation.

Figure 8: Electric fuel-pump design using a flow-type pump as an example
A Fitting cover,
B Electric motor,
C Pump barrel.
1 Electrical connection,
2 Hydraulic connection (fuel outlet),
3 Non-return valve,
4 Pressure-relief valve
5 Commutator with carbon brushes,
6 Motor armature with windings which are energized by the commutator,
7 Magnet,
8 Impeller ring of flow-type pump,
9 Hydraulic connection (fuel inlet).



Electric motor
The plunger-and-barrel assembly (pump element) of the electric fuel pump (Figure 8) is driven by an electric motor. An armature with copper or carbon commutator is used as standard in this engine. Electronic commutating systems without commutators and carbon brushes are increasingly being used in new vehicles that appear on the market. The design of the electric motor depends on the desired delivery quantity for a given system pressure. It is continuously surrounded by flowing fuel and thereby cooled.

End cover
The end cover contains the electrical connection and the pressure-side hydraulic connection. The non-return valve prevents the fuel lines from draining after the electric fuel pump is switched off; this ensures that the system pressure is maintained for a certain period of time after switch-off.

A pressure-relief valve is integrated as required. Most end covers also include the carbon brushes for the commutation system and interference-suppression elements (inductance coils, with capacitors in some applications).

Plunger-and-barrel assembly
The plunger-and-barrel assembly is designed as a positive-displacement or flow-type pump.

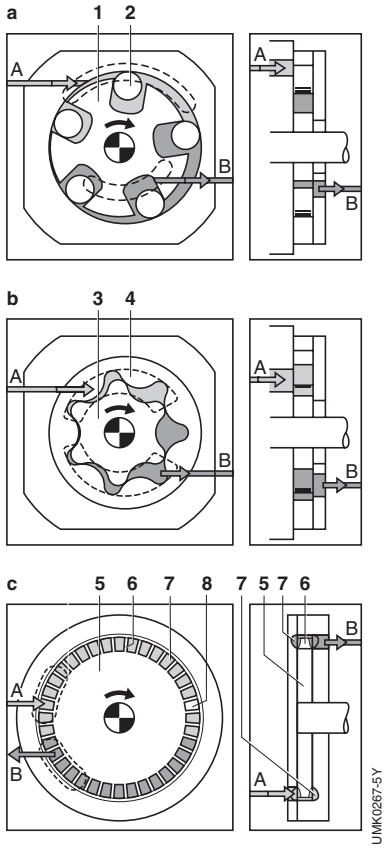
Positive-displacement pump
In a positive-displacement pump, volumes of liquid are basically drawn in and transported in a closed chamber (apart from leaks) by rotation of the pump element to the high-pressure side. A roller-cell pump (Figure 9a), an internal-gear pump (Figure 9b) or a screw-spindle pump may be used for the electric fuel pump.

Roller-cell pump
The slotted rotor arranged eccentrically in the pump housing has metal rollers around its circumferences which are

loosely guided in the slot-shaped openings. The centrifugal force generated by the rotation of the slotted rotor and the fuel pressure press the rollers against the pump housing and the driving flanks of the slots. The rollers operate here as circulating seals. The fuel is delivered in the chamber created between in each case two rollers of the slotted rotor and the pump housing. The pumping effect is created by the fact that, after the inlet opening is sealed, the chamber volume is continuously reduced until the fuel leaves the pump through the outlet opening.

Figure 9: Electric fuel-pump principles
a) Roller-cell pump,
b) Internal-gear pump,
c) Flow-type pump.

A Fuel inlet (suction opening),
B Outlet (pressure side).
1 Slotted rotor (eccentric),
2 Roller,
3 Internal drive wheel,
4 Rotor (eccentric),
5 Impeller,
6 Impeller blades,
7 Passage (peripheral),
8 Gas-discharge orifice.



Internal-gear pump
The internal-gear pump consists of an internal drive wheel which engages an eccentrically arranged outer rotor. This outer rotor has one tooth more than the drive wheel. The tooth flanks that are sealed against each other create during the rotation process in their spaces variable chambers which provide the pumping effect.

Positive-displacement pumps are advantageous when used with high-viscosity media, e.g. cold diesel fuel. They have a good low-voltage characteristic, i.e. they have a relatively "flat" delivery-rate characteristic over the operating voltage. Efficiency can be as high as 40 %. The unavoidable pressure pulses may cause noise; the extent of this problem varies according to the pump's design configuration and mounting location.

Whereas in electronic gasoline-injection systems the positive-displacement pump has to a large extent been superseded by the flow-type pump for the classical electric-fuel-pump requirements, it has gained a new field of application as the pre-supply pump on diesel common-rail systems with their significantly greater pressure requirements and viscosity range.

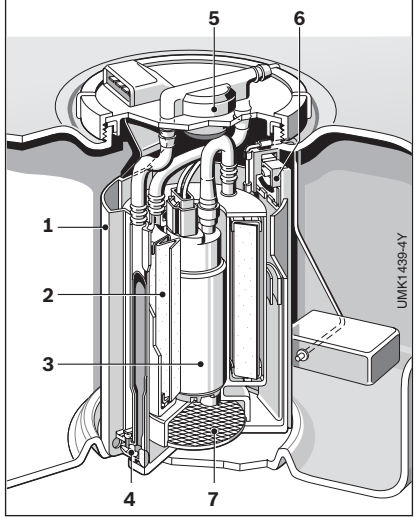
Flow-type pump
The flow-type pump has become the accepted solution for gasoline applications. An impeller ring equipped with numerous peripheral vanes rotates within a chamber consisting of two fixed housing sections (Figure 9c). Each of these sections features a passage along the path of the impeller blades, with openings at one end of the passage starting at the suction openings. From there, they extend to the point where the fuel exits

the plunger-and-barrel assembly at system pressure. A small gas-discharge orifice, located at a specified angular distance from the suction opening, improves performance when pumping hot fuel; this orifice facilitates the discharge of any gas bubbles that may have formed (with minimal leakage).

Pressure builds up along the passage as a result of the exchange of pulses between the impeller blades and the liquid particles. This leads to spiral-shaped rotation of the liquid volume trapped in the impeller and in the passages.

Flow-type pumps feature a low noise level since pressure buildup takes place continuously and is practically pulsation-free. Pump design is also much less complex than the positive-displacement unit. System pressures of up to 650 kPa can be achieved. The efficiency of these pumps can, depending on the operating point, exceed 30 %.

Figure 10: Fuel-supply module
1 Swirl pot,
2 Fuel filter,
3 Electric fuel pump,
4 Suction jet pump,
5 Fuel-pressure regulator,
6 Fuel-level sensor,
7 Prefilter.



Fuel-supply module

Whereas in the early stages of electronic gasoline injection the electric fuel pump was mounted exclusively outside the fuel tank (in-line), it is common practice today to install the pump inside the tank itself. The electric fuel pump forms an integral part of a fuel-supply module (in-tank unit, Figure 10) which may comprise further elements:

– a swirl pot,
– a fuel-level sensor (see Sensors),
– a fuel-pressure regulator for returnless fuel supply systems,
– a prefilter for protecting the pump,
– a pressure-side fine-mesh fuel filter, which will not need replacing for the entire life of the vehicle,
– and electrical and hydraulic connections.
– Tank pressure sensors (for diagnosing tank leaks), fuel-pressure sensors (for demand-regulated systems) and valves for tank ventilation may also be incorporated.

The swirl pot is situated centrally in the fuel tank. The suction side of the electric fuel pump is inserted in this beaker-shaped container. When the tank level is high, fuel flows over the beaker edge, when the level is low, fuel passes through a small opening in the beaker base or through the return line of the electric fuel pump into the swirl pot. When the vehicle is cornering, accelerating and braking, the fuel in the fuel tank flows to one side, which, when the tank level is low, could cause the fuel supply to be interrupted. The fuel can only escape slowly from the swirl pot through the small opening such that the fuel supply is safeguarded for a certain period of time.

The swirl pot can be filled passively, for example by a flap system or a change-over valve, or actively by a suction jet pump. This pump is designed as a venturi nozzle and does not contain any moving parts. The returning fuel flows as a driving medium into this pump and is accelerated after being discharged through the nozzle. In accordance with Bernoulli's law this produces a pressure drop. In this way, fuel from the fuel tank is drawn in through a second inlet and delivered under pressure with the returning fuel to the pump outlet.

High-pressure pumps for gasoline direct injection

The function of the high-pressure pump is to compress the fuel supplied in sufficient quantity by the electric fuel pump at an rpm- and temperature-dependent admission pressure (program-map value) to the level required for high-pressure injection.

Demand-regulated high-pressure pump

Design and operating principle

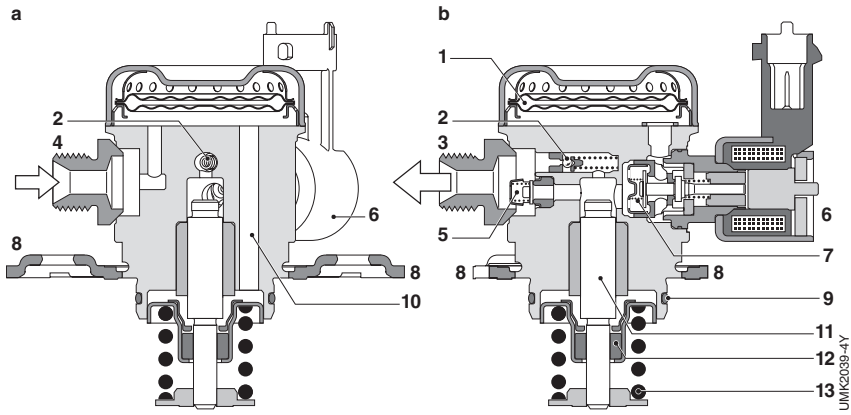
The Bosch demand-regulated high-pressure pump is used in systems for gasoline direct injection for injection pressures up to 20 MPa (2nd-generation gasoline direct injection). Subsequent further development of the second generation increases the pressure range to 25 MPa and improves the acoustics of the fuel-supply control valve. The third generation increases the pressure range to 35 MPa, facilitates a greater delivery quantity, and once again improves the acoustics.

The high-pressure pump is a cam-driven single-barrel pump running in oil (Figure 11) with an integrated fuel-supply control valve (also called the metering unit), an integrated pressure-relief valve on the high-pressure side and an integrated fuel-ressure attenuator on the low-pressure side. To comply with future stringent fuel standards and in response to emission-control legislation, the pump is made of stainless steel and is welded at all the emission-relevant connection points.

The high-pressure pump is mounted as a plug-in pump on the cylinder head. The interface between the engine camshaft and the delivery barrel is a bucket tappet for a double cam (Figure 12a), and a roller tappet for triple and quadruple cams (Figure 12b). This ensures that the cam lift curve is transferred to the delivery plunger (pump plunger) of the high-pressure pump with the requirements with regard to lubrication, Hertzian stress and mass inertia. During the cam lift the tappet follows the contour of the cam. This results in the vertical motion, or stroke, of the delivery plunger. In the delivery stroke the push

Figure 11: Demand-regulated single-cylinder high-pressure pump for 2nd-generation gasoline direct-injection systems

a) View with low-pressure port,
b) View with high-pressure port (on the same plane offset at an angle to low-pressure port).
1 Variable fuel-pressure attenuator, 2 Pressure-relief valve, 3 High-pressure port, 4 Low-pressure port, 5 Outlet valve, 6 Fuel-supply control valve, 7 Inlet valve, 8 Mounting flange, 9 O-ring, 10 Connection to plunger step space (pressure-attenuation function), 11 Delivery plunger, 12 Plunger seal, 13 Plunger spring.



rod absorbs the applied pressure, mass, spring and contact forces. It is rotationally secured in the process.

With a quadruple cam, time synchronization of delivery and injection is possible in a four-cylinder engine, i.e., each injection is also accompanied by a delivery. In this way, it is possible – due to the pressure dip in the rail – on the one hand to reduce excitation of the high-pressure circuit and on the other hand to reduce the rail volume.

In order to ensure that the system pressure can still be sufficiently quickly achieved or varied with the engine at maximum fuel demand, the maximum delivery quantity of the high-pressure pump is configured for maximum demand plus necessary factors which affect delivery behavior (e.g., high-pressure starting, hot gasoline, pump aging, dynamic response).

The volumetric efficiency is derived from the ratio of actually delivered fuel quantity to theoretically possible quantity. This is dependent on the delivery-plunger diameter and stroke. The volumetric efficiency is not constant across all engine speeds. It is dependent on the following factors:

– In the lower speed range: plunger and other leakages,
– In the upper engine speed range: dynamics (inertia and opening pressure) of the inlet and outlet valves,
– In the total speed range: dead volume of the delivery chamber and temperature dependence of fuel compressibility.

Figure 12: High-pressure pump drives
a) Drive via bucket tappet,
b) Drive via roller tappet.
1 Pump plunger, 2 Bucket tappet,
3 Roller tappet, 4 Double cam,
5 Triple cam,
6 Drive shaft (engine camshaft).

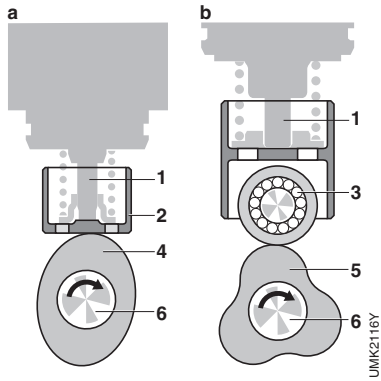
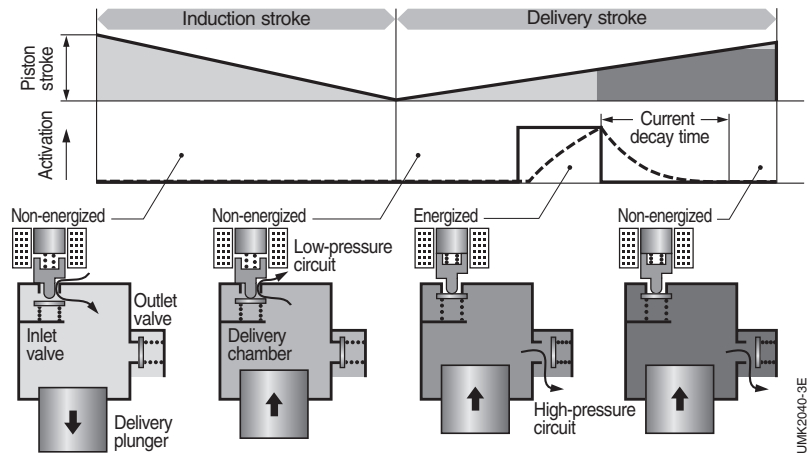


Figure 13: Activation concept of fuel-supply control valve



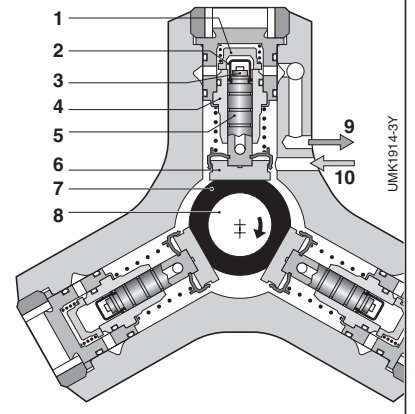
Fuel-supply control valve
Demand control of the high-pressure pump is effected with the fuel-supply control valve (Figure 13). The fuel delivered by the electric fuel pump is drawn into the delivery chamber via the inlet valve of the open fuel-supply control valve. In the subsequent delivery stroke, the fuel-supply control valve remains open after bottom dead center so that fuel that is not needed at the relevant load point is returned at admission pressure to the low-pressure circuit. After the fuel-supply control valve is actuated, the inlet valve closes, the fuel is compressed by the delivery plunger and delivered to the high-pressure circuit. The inlet valve then remains closed during the delivery stroke on account of the pressure in the delivery chamber even when the fuel-supply control valve is non-energized. The engine-management system calculates the time from which the fuel-supply control valve is actuated as a function of delivery quantity and rail pressure. The start of delivery is varied here for demand control.

Fuel-pressure attenuator
The variable fuel-pressure attenuator (Figure 11) integrated in the high-pressure pump, together with the delivery plunger designed as a stepped plunger (backpumping in the induction stroke due to the smaller diameter in the lower area causes a repumping effect), attenuate the pressure pulsations stimulated by the high-pressure pump in the low-pressure circuit and at high rotational speeds ensure a good delivery-chamber charge. The fuel-pressure attenuator takes up the fuel quantity diverted at the relevant operating point via the deformation of its gas-filled metal diaphragms and releases it again in the induction stroke to fill the delivery chamber. Operation with variable admission pressure – i.e., the use of demand-regulated low-pressure systems – is possible here.

Continuous-delivery high-pressure pump
Design and operating principle
The Bosch continuous-delivery high-pressure pump is used in systems for gasoline direct injection for injection pressures up to 12 MPa (1st-generation gasoline direct injection). It is a radial-piston pump with three delivery barrels situated at circumferential offsets of 120° (triple-barrel radial-piston pump, Figure 14).

The delivery quantity of the continuous-delivery high-pressure pump is proportional to the rotational speed. The three barrels deliver fuel at circumferential offsets of 120° in order to ensure overlapping and therefore continuous delivery. This gives rise to minimal pressure pulsations only. This means that, when compared with demand-regulated systems with single-plunger pumps, less demands

Figure 14: Continuous-delivery high-pressure pump for 1st-generation gasoline direct-injection systems (cross-section, schematic)
1 Outlet valve,
2 Displacement space,
3 Inlet valve,
4 Pump barrel,
5 Pump plunger (plunger hollow, fuel inlet),
6 Shoe,
7 Lift ring,
8 Eccentric element,
9 High-pressure port to rail,
10 Fuel supply (low-pressure).



have to be placed on the pump connections and piping. Furthermore, there is no need for a low-pressure attenuator. On the downside, continuous delivery at high pressure means that it has a higher power loss than a demand-regulated pump.

When the pump operates at constant rail pressure or at part load, the pressure of the excess fuel is reduced to admission pressure level by the pressure-control valve mounted on the rail and returned to the intake side of the high-pressure pump. The pressure level in the high-pressure circuit is regulated and adjusted by means of the engine control unit, which specifically actuates the pressure-control valve.

Pressure-control valve
The pressure-control valve (Figure 15) is a proportional control valve which is closed at zero current and actuated by means of a pulse-width-modulated signal. During operation, the energizing of the solenoid coil sets a magnetic force which relieves the load on the spring, lifts the valve ball off the valve seat, and thereby alters the flow cross-section. The pressure-control valve sets the desired rail pressure as a function of the pulse duty factor. The excess fuel delivered by the high-pressure pump is diverted via the drain hole into the low-pressure circuit.

A pressure-limiting function is integrated by way of the valve spring to protect the components against unacceptably high rail pressures, for example, in the event of actuation failure. If one or more of the pump barrels should fail, emergency operation is possible with the intact barrels or by means of the electric fuel pump with admission pressure.

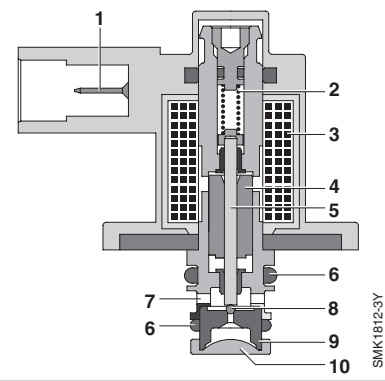
Fuel rail

Intake-manifold injection
The function of the fuel rail is to store the fuel required for injection, to attenuate pulsations and to ensure uniform distribution to all the fuel injectors. The fuel injectors are mounted directly on the rail. Next to the fuel injectors it is possible in systems with fuel return to integrate a fuel-pressure regulator and if necessary a fuel-pressure attenuator in the rail.

Local pressure fluctuations caused by resonance when the fuel injectors open and close is prevented by careful selection of the fuel-rail dimensions. As a result, irregularities in injected fuel quantity which can arise as a function of load and engine speed are avoided.

Gasoline direct injection
The fuel rail (Rail, Figure 16) has the task of storing and distributing the volume of fuel required for each operating point. The fuel is stored by way of its volume and compressibility. In this way, the volume or the inside diameter is engine-power-dependent and must be adapted to the relevant engine demand (injected fuel quantity) and pressure range. Full design of the rail is provided by the dimensioning of the outside diameter or the wall thick-

Figure 15: Pressure-control valve
1 Electrical connection, 2 Valve spring, 3 Solenoid coil, 4 Solenoid armature, 5 Valve needle, 6 Sealing ring (O-rings), 7 Drain hole, 8 Valve ball, 9 Valve seat, 10 Inlet with inlet strainer.



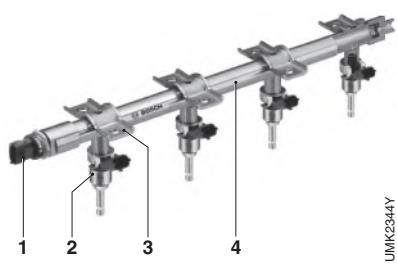
ness of the rail to guarantee strength. The rail is designed to be operationally stable, which is checked by way of special pressure load profiles which have been derived from real operating loads. The volume also ensures hydraulic attenuation in the high-pressure range, i.e., pressure fluctuations in the rail are compensated.

The rail consists of different add-on components which act either as interfaces or as functional groups. Specifically, these are:

– a stainless-steel pipe for accommodating the volume and mounting further attachments (copper-brazed or screwed),
– retainer for securing the rail to the engine block,
– cup for accommodating the fuel injectors,
– high-pressure line connection with integrated restrictor,
– pressure-sensor connection,
– end cap,
– hold-down element for securing the injector position during operation with combustion-chamber pressure and vibration acceleration,
– wiring harness with clipped-on plastic duct (around the pipe).

In part these solid, constructed or deep-drawn attachments are also designed as function modules (e.g., retainer and cup). The fuel injectors are sealed by O-rings, the pressure sensor and the high-pressure connection by a screwed ball-taper geometry.

Figure 16: Rail for gasoline direct injection
1 High-pressure sensor, 2 Fuel injectors, 3 Retainer, 4 Rail.



A particular feature is the “suspended design” of the injector mounting, which ensures a low load thanks to its pressure-compensated construction. The drawback of this variant is the increased amount of space required.

The biggest challenge in the development of rails is the large number of necessary solutions in that it must always be integrated in the engine in an already specified installation space.

The rail for 1st-generation fuel-injection systems is designed for a pressure range up to 12 MPa (plus 0.5 MPa pressure-relief valve opening pressure). For the 2nd generation the pressure range extends up to 25 MPa (plus 5.0 MPa pressure-relief valve opening pressure), for the 3rd generation according to the high-pressure pump up to 35 MPa. The acoustics of the rail in the overall system are also improved as part of the further development.

Fuel-pressure regulator

Function
With intake-manifold injection the amount of fuel injected by the fuel injector depends on the injection period and the pressure differential between the fuel pressure in the fuel rail and the back pressure in the intake manifold. On fuel systems with fuel return, the influence of pressure is compensated for by a fuel-pressure regulator which maintains the differential between fuel pressure and intake-manifold pressure at a constant level. This pressure regulator permits just enough fuel to return to the fuel tank so that the pressure drop across the fuel injectors remains constant. In order to ensure that the fuel rail is efficiently flushed, the fuel-pressure regulator is normally located at the end of the rail.

On returnless fuel systems, the fuel-pressure regulator is part of the fuel-supply module in the fuel tank. The fuel-rail pressure is maintained at a constant level with reference to the ambient pressure. This means that the pressure differential between fuel-rail pressure and intake-manifold pressure is not constant and must be taken into account when the injection duration is calculated.

Mounting on the rail
The fuel-pressure regulator (Figure 17) is of the diaphragm-controlled overflow type. A rubber-fabric diaphragm divides the pressure regulator into a fuel chamber and a spring chamber. Through a valve holder integrated in the diaphragm, the compression spring forces a movable valve plate against the valve seat. As soon as the pressure applied to the diaphragm by the fuel exceeds the spring force, the valve opens again and permits just enough fuel to flow back to the fuel tank that equilibrium of forces is achieved again at the diaphragm.

The spring chamber is pneumatically connected with the intake manifold behind the throttle valve. As a result, the intake-manifold vacuum also has an effect in the spring chamber. There is therefore the same pressure ratio at the diaphragm as at the fuel injectors. This means that the pressure drop across the fuel injectors is solely a function of spring force and diaphragm surface area, and therefore remains constant.

Figure 17: Fuel-pressure regulator for mounting on the rail

1 Intake-manifold port (omitted for application in the fuel-supply module),
2 Compression spring,
3 Spring chamber,
4 Valve holder,
5 Rubber-fabric diaphragm,
6 Valve,
7 Nozzle,
8 Fuel feed to fuel chamber,
9 Fuel return.

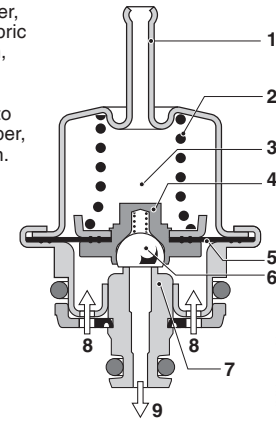
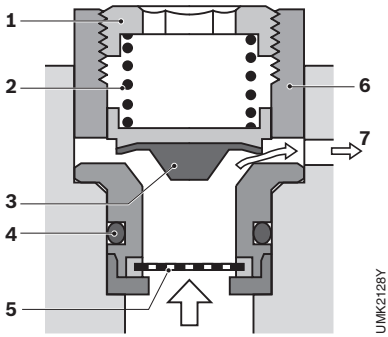


Figure 18: Fuel-pressure regulator for applications in the fuel-supply module
1 Adjusting screw (for primary pressure),
2 Compression spring,
3 Valve plate,
4 O-ring,
5 Prefilter,
6 Housing,
7 Discharge to swirl pot.



Installation in the fuel-supply module
For applications in the fuel-supply module the fuel-pressure regulator can have a simpler design that the pressure regulator mounted on the rail. These simple pressure regulators (Figure 18) have therefore become the accepted choice. The fuel flows through a coarse prefilter (protecting the pressure regulator against fouling) against the valve plate. An annular gap opens when the fuel pressure exceeds the spring force that presses the valve plate against the seal seat. The fuel flows through this annular gap and returns through the pressure regulator housing directly to the swirl pot.

Fuel-pressure attenuator

Function, design and operating principle
The repeated opening and closing of the fuel injectors, high-pressure pump pulsations (the principal cause today) and the periodic supply of fuel when electric positive-displacement fuel pumps are used lead to fuel-pressure oscillations. These can cause pressure resonances which adversely affect fuel-metering accuracy. It is even possible that under certain circumstances noise can be caused by these vibrations being transferred to the fuel tank and the vehicle bodywork through the mounting elements of the electric fuel pump, fuel lines, and fuel rail. These problems are alleviated by the use of special-design mounting elements and fuel-pressure attenuators.

The fuel-pressure attenuator has a similar design to the fuel-pressure regulator. Here too, a spring-loaded diaphragm separates the fuel chamber from the air chamber. The spring force is set in such a way that the membrane rises from its seat as soon as the fuel delivery pressure reaches its working range. This means that the fuel chamber is variable and not only absorbs fuel when pressure peaks occur, but also releases fuel when the pressure drops. In order to always operate in the most favorable range when the absolute fuel pressure fluctuates due to conditions at the manifold, the spring chamber can be provided with an intake-manifold connection.

Similar to the fuel-pressure regulator, the fuel-pressure attenuator can also be attached to the rail or installed in the fuel line. In the case of gasoline direct injection, it can also be attached to the high-pressure pump.

Mixture formation

Basic principles

A/F (Air/fuel) mixture
To be able to operate, a spark-ignition engine requires an air-fuel mixture. The mixture for ideal, theoretically complete combustion requires a mass ratio of 14.7:1 (stoichiometric ratio), i.e. 14.7 kg of air are required to burn 1 kg of fuel. Or: 1 *l* of fuel burns completely in roughly 9,500 *l* of air.

The specific fuel consumption is essentially dependent on the air-fuel ratio. Excess air is required in order to ensure genuine complete combustion, and thus as low a fuel consumption as possible. However, limits are imposed due to the flammability of the mixture and the available combustion time.

The A/F mixture also has a decisive impact on the efficiency of exhaust-gas treatment systems. State-of-the-art technology is represented by the three-way catalytic converter. However, it needs a specific stoichiometric A/F ratio in order to operate at maximum efficiency. This can reduce damaging exhaust gas components by higher than 99 %. The homogeneous engines available today are therefore operated with a stoichiometric mixture as soon as their operating status allows this. By way of departure, individual direct-injection engines are also operated with a lean mixture (stratified-charge operation) in the interest of fuel economy.

Certain engine operating statuses require mixture adaptation. Selective modifications to the mixture composition are required e.g. in the case of a cold engine.

Excess-air factor λ
The excess-air factor or air ratio λ (lambda) has been chosen to designate the extent to which the actual air/fuel mixture differs from the theoretically necessary mass ratio (stoichiometric ratio 14.7:1). λ denotes the ratio of air mass in the cylinder to the air requirement in the case of stoichiometric combustion.

Stoichiometric ratio
$\lambda = 1$: The inducted air mass corresponds to the theoretically required air mass.

Air deficiency
$\lambda < 1$: This indicates air deficiency and therefore a rich A/F mixture. The maximum engine output results at $\lambda = 0.85$ to 0.95.

Enriching the air-fuel mixture may be necessary to protect components against thermal overload (component protection). This reduces the exhaust-gas temperature, since the fuel is converted not completely to CO_2, but to CO. The enthalpy for the reaction to CO is lower than for complete conversion to CO_2. A lower thermal load e.g. for exhaust-gas turbocharger, λ sensor and catalytic converter results from the lower exhaust-gas temperature.

Figure 1: Effects of excess-air factor λ on specific fuel consumption b_e and uneven running at constant power output

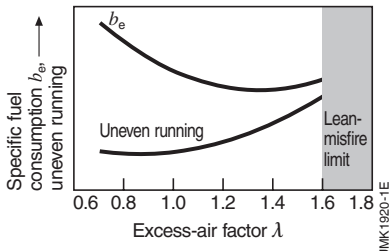
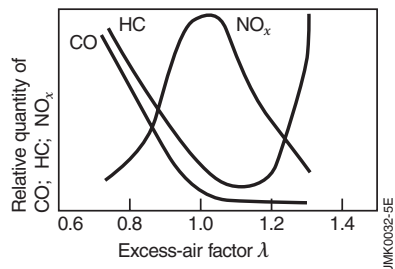


Figure 2: Effect of excess-air factor λ on pollutant composition



However, this also entails increased consumption. Enrichment is therefore only performed for high-load points for which an exhaust-gas temperature of over 900 °C would be reached.

Air surplus
$\lambda > 1$: In this range, there is an air surplus and thus a lean mixture.

This excess-air factor is characterized by reduced fuel consumption (Figure 1), but also by reduced power output. The maximum value for λ – the "lean-misfire limit" – that can be achieved is very strongly dependent on the construction of the engine and the mixture-formation system used. The mixture is no longer ignitable at the lean-misfire limit because of the lower number of HC molecules. Combustion misses occur, and this is accompanied by a marked increase in uneven running.

Influence of λ on fuel consumption and exhaust-gas composition
Spark-ignition engines achieve the lowest fuel consumption at constant engine output dependent on the engine at 20 to 50 % air surplus (λ = 1.2 to 1.5). This potential can however only be achieved with lean-burn concepts (homogeneous-lean or stratified-charge operation).

For a typical engine with intake-manifold injection, Figures 1 and 2 show the dependency of the specific fuel consumption and uneven running, as well as development of pollutants, on the excess-air factor at constant engine output. It can be deduced from these graphs that there is no ideal excess-air factor at which all the factors assume the most favorable value. In order to implement "optimal" consumption at "optimal" power output, excess-air factors of λ = 0.95 to 1.05 have proven conducive to the achievement of objectives for engines with intake-manifold injection and for homogeneous direct-injection engines – above all in view of exhaust-gas treatment with the three-way catalytic converter, which can optimally convert all three pollutant components only in a narrow range around λ = 1.

To achieve this, the inducted air mass is recorded precisely and an exactly metered fuel mass is apportioned by means of λ regulation in the engine-management system.

Intake-manifold injection
For optimum combustion in engines with intake-manifold injection, not only is a precise injected fuel quantity necessary, but also a homogeneous A/F mixture. This requires efficient fuel atomization. If this precondition is not satisfied, large fuel droplets will precipitate on the intake manifold or on the combustion-chamber walls. These large droplets cannot fully combust and will result in increased hydrocarbon emissions (HC).

Gasoline direct injection
Direct-injection engines with charge stratification involve different combustion conditions so that the lean-misfire limit occurs at significantly higher global λ values. In the part-load range, these engines can therefore be operated with considerably higher excess-air factors (up to λ = 4). Here too the area around the spark plug requires locally a stoichiometric mixture for good flammability.

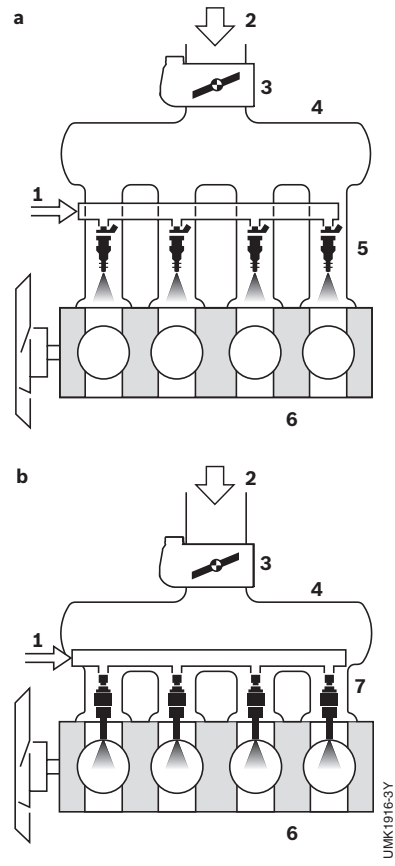
Unfortunately, in this case of a globally lean mixture, only two of the three ways of the three-way catalytic converter function. CO emissions and unburnt hydrocarbons can still be oxidized since sufficient oxygen is available. However, the third chief reaction, reducing the NO_x emissions, cannot be carried out sufficiently in lean-operation mode, since insufficient reduction partners are available. For this reason, in addition to catalytic exhaust-gas treatment with a three-way catalytic converter, a further exhaust-gas treatment facility must be installed for NO_x reduction (e.g. an NO_x storage catalytic converter).

Mixture-formation systems

It is the job of fuel-injection systems, or carburetors, to provide an A/F mixture which is adapted as well as possible to the relevant engine operating status. Without fuel-injection systems, particularly electronic systems, it would not be possible to maintain the increasingly narrowly defined limits for mixture composition and thus to adhere to the stipulations of the emission-control legislation. These requirements and the significant improvements in fuel metering and mixture formation together with the benefits in terms of fuel consumption, driving performance and power output have meant that, in the automotive sector, gasoline injection has completely superseded the carburetor in moder engines.

Until the beginning of this century, the automotive industry almost exclusively used systems in which A/F mixture formation takes place outside the combustion chamber (intake-manifold injection, Figure 3a). Systems with internal A/F mixture formation, i.e. where the fuel is injected directly into the combustion chamber (gasoline direct injection, Figure 3b), are designed to reduce fuel consumption even further and to increase power output, and have therefore become increasingly more important.

Figure 3: Schematic representation of fuel-injection systems
a) Intake-manifold injection,
b) Gasoline direct injection.
1 Fuel,
2 Air,
3 Throttle device,
4 Intake manifold,
5 Fuel injectors,
6 Engine,
7 High-pressure fuel injectors.



Intake-manifold injection

Gasoline injection systems for external mixture formation are characterized by the fact that the A/F mixture is created outside of the combustion chamber (in the intake manifold). The state of the art is represented by electronic intake-manifold injection systems where the fuel is injected cyclically for each individual cylinder directly ahead of the intake valves (Figure 4).

Systems that are based on mechanical-continuous fuel injection (K-Jetronic) or a central fuel injection arranged upstream of the throttle valve (Mono-Jetronic) are no longer of any significance for new developments.

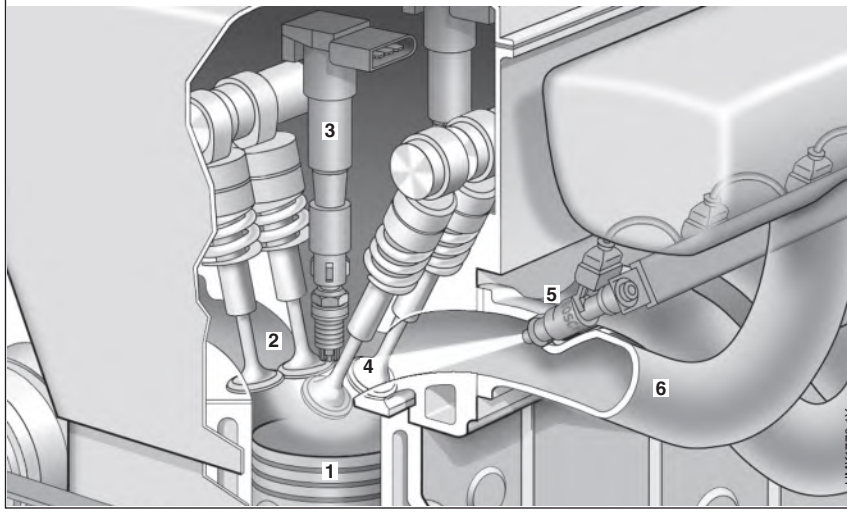
Requirements

The high requirements for engine smooth-running and exhaust emissions make high demands on the A/F mixture composition in each power cycle. Precise injection timing is vital, as well as precise metering of the injected fuel mass as a factor of the engine intake air. In electronic multipoint fuel-injection systems, therefore, not only is each engine cylinder assigned an electromagnetic fuel injector, but this fuel injector is also activated individually for each cylinder. The engine control unit has the task of calculating both the required fuel mass for each cylinder and the correct start of injection for the fuel mass drawn in and the current engine operating status. The injection time required to inject the calculated fuel mass is a function of the opening cross-section of the fuel injector and the pressure differential between the intake manifold and the fuel supply system.

Fuel system

In intake-manifold injection systems, fuel is sent via the electric fuel pump (see Fuel supply), the fuel supply lines, and the fuel filter at typical system pressures of 3 to 7 bar to the fuel rail, which ensures that the fuel is evenly distributed to the fuel injectors. How the fuel is prepared by the fuel injectors is extremely important for the quality of the A/F mixture. It is essential that the fuel is atomized into very fine droplets. The spray shape and spray dispersal angle of the fuel injectors are adapted to the geometric shape of the intake manifold and cylinder head (see Intake-manifold fuel injector).

Figure 4: Principle of intake-manifold injection
1 Cylinder with piston, 2 Exhaust valves, 3 Ignition coil with spark plug, 4 Inlet valves, 5 Fuel injector, 6 Intake manifold.



Fuel injection

The exactly metered fuel mass is injected directly upstream of the cylinder intake valve(s). As the air enters through the throttle valve there is a pulse exchange with the finely atomized fuel, and the droplets on the surface mix increasingly with the air and therefore evaporate to a large extent (Figure 5). The required air-fuel mixture can therefore be formed at the right time by varying the point of injection.

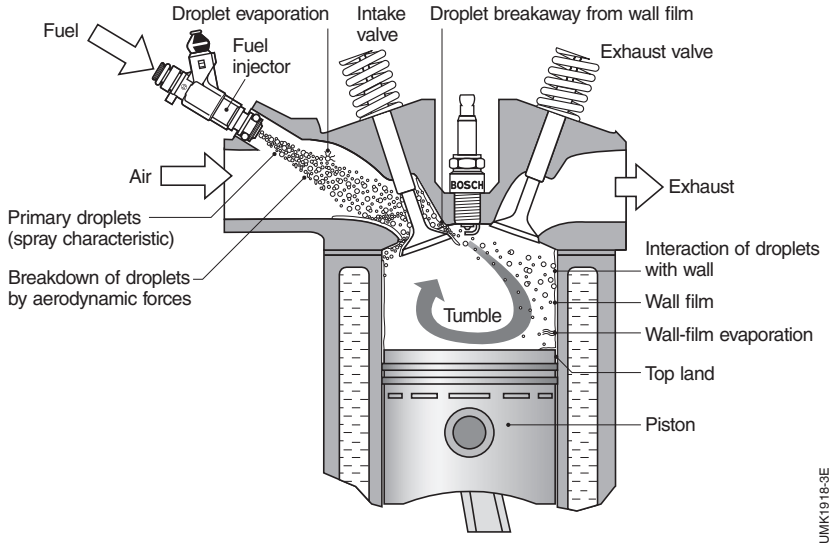
The amount of time available for mixture formation can be increased by injecting the fuel into the intake valves that are still closed. How much air-fuel mixture is created directly after injection depends on the quality of the fuel injector (primary droplet size). In this case, the primary evaporation occurs only when the intake valves open in response to the high flow velocities with an initially very small valve gap.

A proportion of the fuel precipitates on the intake duct (part of the intake manifold leading directly to the combustion chamber) and on the lower edge of the intake valves as wall film; the thickness of this film essentially depends on the pressure in the intake manifold, and thus on the engine load condition. In the case of non-stationary (dynamic) engine operation, this precipitation can lead to temporary deviations in the desired λ value ($\lambda = 1$), which means that the fuel mass stored in the wall film is to be kept as low as possible.

Wall-coating effects in the intake duct, particularly in cold-start conditions, cannot be neglected either. As fuel does not evaporate sufficiently, more fuel is required initially in the starting phase in order to create an ignitable A/F mixture. When the intake-manifold pressure then drops, parts of the previously formed wall film will vaporize. This may result in increased HC emissions if the catalytic converter is not running at operating temperature.

Irregular fuel injection may also result in the formation of wall films in the combustion chamber, and may in turn become critical emission sources. Defining the geometric alignment of fuel sprays ("spray targeting") will allow the selection of suit-

Figure 5: Mechanisms and factors influencing mixture formation in intake-manifold injection



able fuel injectors which will control and minimize manifold-wall fuel condensation in the area of the intake duct and the intake valves.

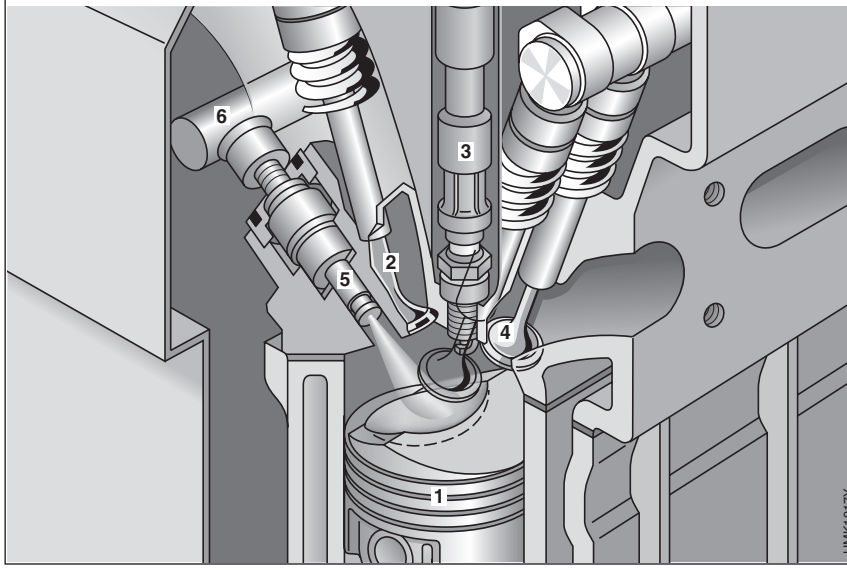
The intake manifolds used can be optimally adapted to the combustion air flow and the dynamic gas requirements of the engine.

Gasoline direct injection

In gasoline direct injection (GDI), unlike intake-manifold injection, pure air flows through the intake valves into the combustion chamber. Only then is the fuel injected into the air via an injector (see high-pressure fuel injector) located directly in the cylinder head (internal mixture formation, Figure 6). There are basically two main operating modes. Fuel injection in the intake stroke is called homogeneous operation, while fuel injection during compression is called stratified-charge operation. There are also various special modes, which are either a mixture of the two main operating modes or a slight variation of them.

In stratified-charge operation, the air is not restricted; the A/F mixture is lean. The excess air in the exhaust gas prevents conversion of the nitrogen oxides by means of a three-way catalytic converter. These direct-injection systems therefore require exhaust-gas treatment with an additional NO_x storage catalytic converter. For this reason, mostly direct-

Figure 6: Principle of gasoline direct injection
1 Cylinder with piston, 2 Intake valves, 3 Ignition coil with spark plug, 4 Exhaust valves, 5 High-pressure fuel injector, 6 Rail.



injection systems that work exclusively in the homogeneous mode are currently being placed on the market.

Homogeneous operation

In homogeneous operation, mixture formation is similar to intake-manifold injection. The mixture is formed in a stoichiometric ratio (λ = 1). However, from a mixture formation point of view, there are some differences. For instance, there is no flow process around the intake valve to promote mixture formation, and there is much less time available for the mixture formation process itself. Whereas injection can take place across the overall 720° crankshaft (stored and synchronous with induction) of the four operating cycles in the case of intake-manifold injection, only an injection window of 180° crankshaft remains in the case of gasoline direct injection.

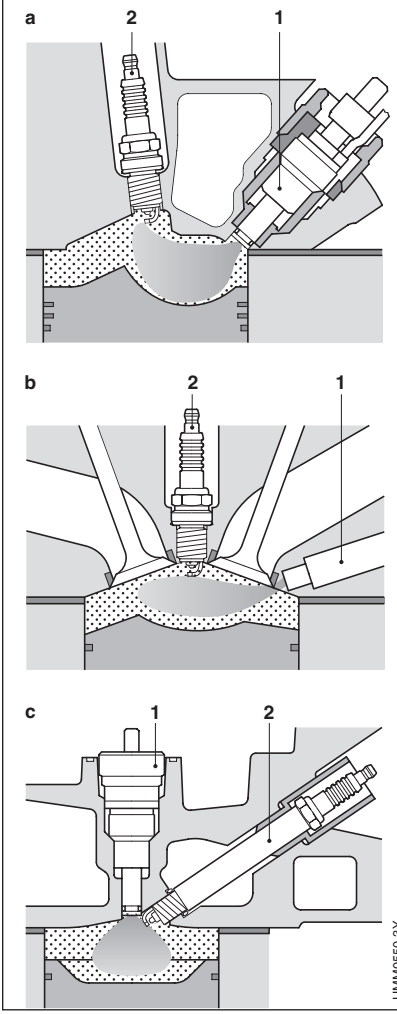
Fuel injection is only permitted in the induction stroke. This is because, prior to this, the exhaust valves are open and unburned fuel would otherwise escape into the exhaust-gas train. This would cause high HC emissions and catalytic converter problems. In order to deliver a sufficient quantity of fuel in this shortened period, the fuel flow through the injector must be increased for gasoline direct injection. This is achieved mainly by increasing fuel pressure. The increase in pressure brings with it an additional advantage as it increases turbulence in the combustion chamber, which in turn promotes mixture formation. The fuel and air can therefore be completely mixed, even though the air/fuel interaction time is shorter compared with intake-manifold injection.

Stratified-charge operation

In stratified-charge operation, a distinction is made between several combustion strategies. All strategies have one thing in common, namely they all attempt to achieve charge stratification. This means that, instead of supplying the corresponding stoichiometric air flow rate to the fuel quantity required for a specific load point by adjusting the throttle valve, the full air flow rate is supplied. Only a portion of the air interacts with the fuel, while the rest of the fresh air surrounds this stratified-charge cloud. This dethrottling action facilitates, together with the cooling effects of the directly injected fuel (evaporative emis-

Figure 7: Combustion processes for gasoline direct injection

a) Wall-directed combustion,
b) Air-directed combustion,
c) Spray-directed combustion.
1 High-pressure fuel injector,
2 Spark plug.



sion), an increase in compression and thus considerable fuel-saving potential.

Wall-directed combustion process
In a wall-directed combustion process, fuel is injected into the combustion chamber from the side (Figure 7a). A recess in the piston crown deflects the fuel spray in the direction of the spark plug. Mixture formation takes place on the path from the injector tip to the spark plug. As the mixture-formation time is very short during injection in the compression stroke (stratified-charge operation), the fuel pressure for this system must often be even higher than for homogeneous operation. The increased fuel pressure shortens the injection time and increases interaction with the air because pulse reflection is greater.

The disadvantage of the wall-directed combustion process is that fuel condenses on the piston wall, which increases HC emissions. As mixture-formation time is short, the stratified-charge cloud usually contains rich mixture zones at higher loads, and this increases the risk of soot production. At low loads, the pulse of the fuel droplets, which is used as a means of transporting the stratified-charge cloud to the spark plug, is low due to the small fuel mass. As a result, the flow must usually be restricted here so that the fuel meets with a lower density of air and thus a lower air resistance.

Air-directed combustion process
In principle, an air-directed combustion process works in exactly the same way as a wall-directed process. The main difference is that the stratified-charge cloud does not interact directly with the piston recess. Instead, it moves on a cushion of air (Figure 7b). The cushion of air is generated by the air present in the cylinder. In contrast to the wall-directed combustion process, with air-directed combustion the fuel-spray injection angle is much flatter such that the fuel spray cannot penetrate the full amount of air up to the piston. This solves the problem of fuel condensing on the piston recess.

Compared with wall-directed combustion processes, the air flows in air-directed combustion processes cannot be fully reproduced. This gives rise to large variations from injection to injection and thus frequently to poor combustion stability right through to individual misfires.

Often, the actual combustion processes are a mixture of wall-directed and air-directed processes, depending on the operating point in each case.

Spray-directed combustion process
The spray-directed combustion process is visually different from the other two processes in that the injector is installed at a different location. It is located top center and injects vertically down into the combustion chamber (Figure 7c). The fuel spray is not deflected. As a result of the high injection pressure (200 bar), an air-fuel mixture forms around the fuel spray already during the injection process. Located directly next to the injector is the spark plug, which ignites the air-fuel mixture.

As a result, with the spray-directed combustion process the mixture-formation time is very short. This requires an even higher fuel pressure for the spray-directed combustion process. This combustion process can eliminate the disadvantages of fuel condensing on the manifold walls, air-flow dependency, and flow restriction at low loads. It therefore has the greatest potential for fuel saving. Nevertheless, the short mixture formation time is a huge challenge for the fuel-injection and ignition systems.

Other operating modes
In addition to homogeneous and stratified-charge operation, there are also special operating modes. They include “changeover between operating modes” (homogeneous-stratified mode), “catalytic-converter heating”, and “knock protection” (homogeneous-split mode), and “homogeneous-lean mode”. These operating modes are sometimes very complex and only ever represent short, temporary states. Therefore they are frequently not operating modes for longer operation.

Fuel injectors

Essentially, the mixture-formation components must ensure that the air-fuel mixture is formed in the most optimal way possible to facilitate combustion. In intake-manifold injection, this is mainly the task of the fuel injector, while in gasoline direct injection, the high-pressure fuel injector can be assisted by a turbulence flap.

Essentially, there are two different ways of actuating the fuel injectors. If one injector is operated by means of a conventional electromagnet, the other has a piezoelectric stack serving as the actuator.

Intake-manifold fuel injector

Design and function

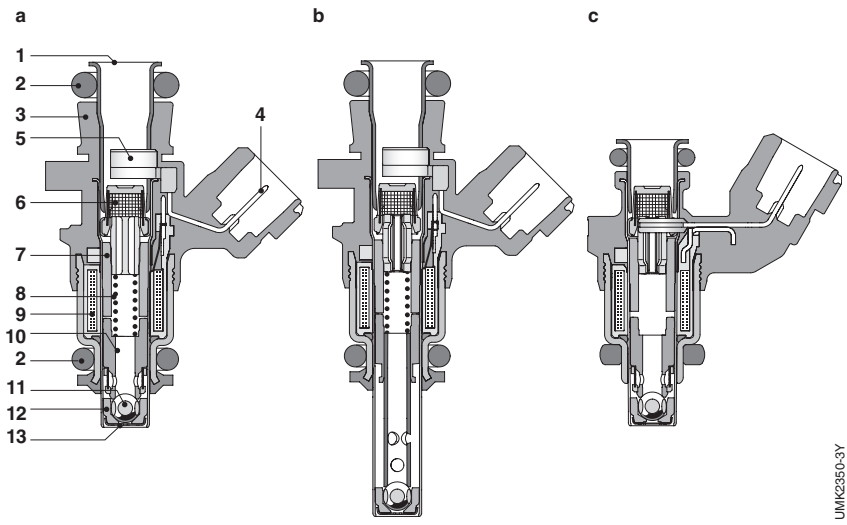
Essentially, electromagnetic fuel injectors are comprised of the following components: the valve housing with electrical connection and hydraulic port, the solenoid coil, the moving valve needle with solenoid armature and valve ball, the valve seat with injection-orifice plate, and the valve spring.

Figure 8a and Figure 8b show the EV14 fuel injector, as is used in intake-manifold injection. The new reduced-size fuel-injector variant EV-SE (Figure 8c) has been developed for small installation spaces and small engines (e.g. for two-wheeled vehicles). Both injector variants are also available with an extended injection point.

To adapt the installation position optimally to the intake-manifold and intake-port geometries, different EV14 design variants are available to choose from: Three overall lengths are available with different electrical connector variants and optionally with an extended injection point (Figure 9.)

A filter strainer in the fuel inlet protects the fuel injector against contamination. The sealing ring (O-ring) on the hydraulic port seals off the injector at the fuel rail. The lower sealing ring provides the seal between the injector and the intake manifold.

Figure 8: Fuel injectors (Bosch) for intake-manifold injection
a) Fuel injector, standard version EV-14fT (flat tip),
b) Fuel injector EV14-xT (extended tip) with extended injection point,
c) Fuel injector EV-SE of reduced size for use e.g. in motorcycles.
1 Hydraulic port, 2 Sealing ring (O-ring), 3 Valve housing, 4 Electrical connection, 5 Connector clip, 6 Filter strainer, 7 Internal pole (stop for solenoid armature), 8 Valve spring, 9 Solenoid coil, 10 Valve needle with solenoid armature, 11 Valve ball, 12 Valve seat, 13 Injection-orifice plate.



When the solenoid coil is de-energized, the valve needle and valve ball are pressed against the cone-shaped valve seat by the valve spring and the force exerted by the fuel pressure. The fuel-supply system is thus sealed off from the intake manifold.

When the solenoid coil is energized, this generates a magnetic field which attracts valve-needle solenoid armature. The valve ball lifts off the valve seat and the fuel is injected. When the excitation current is switched off, the valve needle closes again due to spring force.

The fuel is atomized by an injection-orifice plate. The orifices are stamped out of the plate and ensure that the injected fuel quantity remains highly reproducible. The spray pattern and the atomization quality of the fuel leaving the injector is produced by the number of orifices and their configuration (maximum twelve).

The injected fuel quantity per unit of time is determined mainly by the system pressure in the fuel-supply system (typically 3 to 4 bar, corresponding to 300 to 400 kPa), the back pressure in the intake manifold, and the geometry of the fuel-exit area.

Spray formation and spray targeting

An injector's spray formation, i.e. its spray shape, spray angle, and fuel-droplet size, influences the formation of the air/fuel mixture. Individual geometries of intake manifold and cylinder head make it necessary to have different types of spray formation. Different spray-formation variants are available in order to satisfy with these requirements.

Tapered spray

Individual fuel sprays emerge through the openings in the injection-orifice plate (Figure 10a). These fuel sprays combine to form a spray taper.

Tapered-spray injectors are typically used in engines which feature only one intake valve per cylinder.

Dual spray

Dual-spray formation (Figure 10b) is as a rule used in engines which feature two intake valves per cylinder. The holes in the injection-orifice plate are arranged in such a way that two fuel sprays – which are each formed from a number of individual sprays that produce a homogeneous spray shortly after leaving the injection or-

Figure 9: Three overall lengths for the EV14 fuel injector, optionally with extended injection point

a) EV14, fuel injector, compact, extended injection point,
b) EV14, fuel injector, compact,
c) EV14, fuel injector, standard,
d) EV14, fuel injector, long.

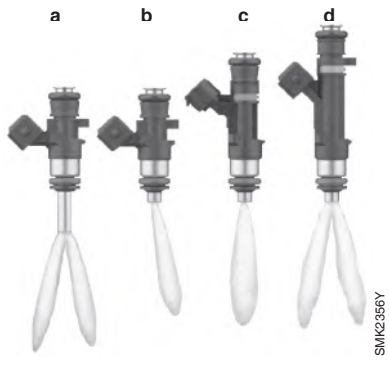
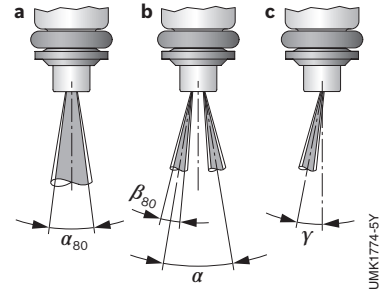


Figure 10: Spray geometries

a) Tapered spray,
b) Dual spray,
c) Gamma angle.
α_{80} 80 % of fuel is inside the angle α,
α Angle between two individual sprays,
β_{80} 80 % of fuel is inside the angle β,
γ Offset spray angle.

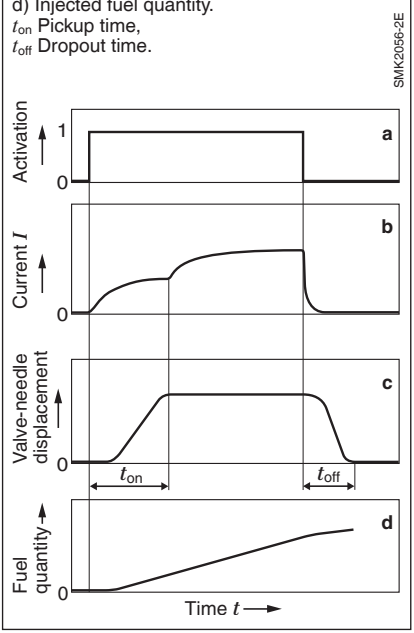


ifices – impact ahead of the intake valves or against the web between the intake valves.

Gamma angle
Compared to the injector's principle axis, the fuel spray in this case (single spray and dual spray) is at an angle: the offset spray angle (Figure 10c).

Injectors with this spray shape are mostly used when installation conditions are difficult, for example when the injectors are installed so steeply in the intake manifold that the required spray targeting can only be achieved by the use of a gamma-angle spray.

Figure 11: Activation of fuel injector
a) Activation signal,
b) Current curve,
c) Valve-needle lift,
d) Injected fuel quantity.
t_{on} Pickup time,
t_{off} Dropout time.



Electrical activation
The output module in the engine ECU actuates the injector with a switching signal (Figure 11a). The current in the solenoid coil rises (Figure 11b) and causes the valve needle to lift (Figure 11c). After the pickup time has elapsed, the maximum valve-needle lift is achieved at the solenoid-armature stop at the internal pole. Depending on the solenoid-valve design, the current continues to rise after the armature stop because the magnetic circuit at this point is not yet at 100 % saturation. The kink in the current curve is triggered by the armature stop.

Fuel is sprayed as soon as the valve ball lifts off its seat. The quantity of fuel injected during an injection pulse is shown in Figure 11d.

Because the magnetic field is not decreased abruptly after activation is shut down, the valve closes with a delay. The valve is fully closed again after the dropout time has elapsed.

The non-linearity during the valve pickup and dropout phases must be compensated for throughout the period that the injector is activated (injection duration). The speed at which the valve needle lifts off its seat is also dependent on the battery voltage. Battery-voltage-dependent injection-duration extension corrects these influences.

Electromagnetic high-pressure fuel injector

The high fuel pressure required for gasoline direct injection (nominal pressure up to approximately 35 MPa) and the installation location give rise to extended requirements for the fuel-injector component when compared with intake-manifold fuel injectors.

Design and function
The function of the high-pressure fuel injector is to meter and atomize the fuel. Atomization ensures that the fuel is quickly mixed with the air in the combustion chamber. In the process, the A/F mixture is formed within a physically restricted area.

The high-pressure fuel injector (Figure 12) consists of the individual components:

– housing with hydraulic port and electrical connection, tolerance compensation and acoustic decoupling element, and a locking element,
– valve seat with injection orifices and pre-stages,
– valve needle with mechanically isolated solenoid armature and spring,
– return spring with setting sleeve,
– coil with conductor bar.

Depending on the installation position in the cylinder head – lateral or central – either short or long valves are used. The long valve is typically based on a short valve which is adapted to the required length with an extension tube for the hydraulic port and an extended conductor bar for the electrical connector. The current generation has an valve-tip diameter reduced to 6 mm to satisfy the requirements with regard to installation space in downsizing engines.

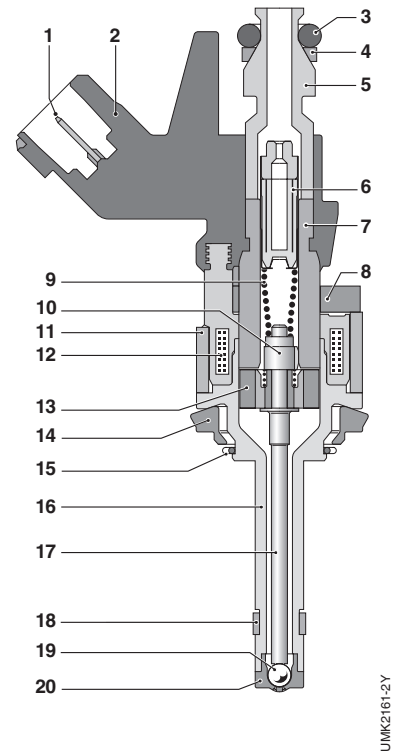
When current flows through the coil, it generates a magnetic field in conjunction with the other components of the magnetic circuit. This magnetic field accelerates the solenoid armature when the valve needle is still at rest. When the solenoid armature contacts the upper stop, the latter lifts the valve needle against spring force and the hydraulic load off the valve seat and opens the valve outlet opening. As the fuel pressure is considerably higher than the combustion-chamber pressure, the fuel is injected into the combustion chamber.

When the current is switched off, the valve needle and the hydraulic load on the valve needle are pushed back into the seat by the return spring, and this stops the injection process. The solenoid armature is returned by the spring to its rest position on the lower stop.

Reproducible fuel quantities can be metered by defining the opening and keeping the opening cross-section constant with the needle fully raised. Here, the fuel quantity is dependent on the pressure in the fuel rail, the backpressure reaction in the combustion chamber, and the valve opening time.

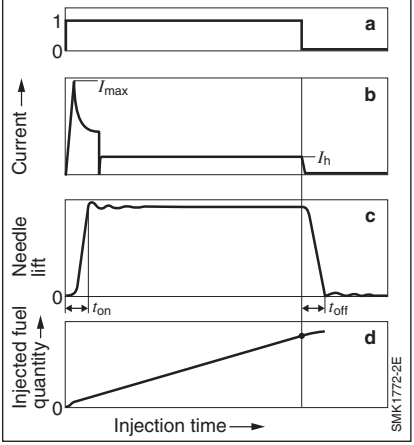
Figure 12: Electromagnetic high-pressure fuel injector (Bosch) for gasoline direct injection

1 Electrical connection,
2 Extrusion coating,
3 O-ring, 4 Support plate,
5 Conductor barrel,
6 Filter with setting sleeve,
7 Internal pole,
8 Cover,
9 Compression spring,
10 Stop ring,
11 Magnetic sleeve,
12 Solenoid coil,
13 Solenoid armature,
14 Support element,
15 Snap ring,
16 Valve housing,
17 Valve needle,
18 Sealing ring,
19 Valve ball,
20 Valve seat with injection orifices.



Spray formation
Predominantly multi-hole valves are used to prepare the fuel. By suitable spray configuration the injection-orifice alignment is adapted to match the installation position to the respective combustion chamber. Here the sprays must be configured in such a way that fuel and fresh air mix thoroughly in the combustion chamber and in the process valves, pistons and combustion-chamber walls are moistened with fuel as little as possible. The fuel injectors are typically configured with five to seven sprays, depending on the desired throughflow, which is configured to match the output per cylinder and thereby determines the mean orifice diameter. The penetration depth of the sprays can be reduced for individual injection orifices by a diameter that is smaller than the mean diameter of all the injection orifices. This makes it possible to reduce the moistening of the piston and the combustion-chamber walls with liquid fuel. Other sprays with a larger orifice diameter are provided to achieve the required throughflow. The injection orifice is for the most part supplemented by a pre-stage. The purpose of this pre-stage is to keep the injection orifice free from deposits which are produced depending on the boundary operating and environmental conditions and on the fuel used.

Figure 13: Signal curves for activating the high-pressure fuel injector
a) Triggering signal calculated by the ECU,
b) Current curve in the high-pressure fuel injector,
c) Valve-needle lift,
d) Injected fuel quantity.
I_{max} Maximum current in the booster phase,
I_h Holding current,
t_{on} Pickup time,
t_{off} Dropout time.



Electrical activation
To ensure a defined and reproducible injection process, the high-pressure fuel injector is activated by a complex current curve. The engine control unit delivers a digital signal for this purpose. An output module uses this signal (Figure 13a) to generate the actuating signal (Figure 13b) for the high-pressure fuel injector.

A DC/DC converter in the engine ECU generates the booster voltage of 65 V. This voltage creates in conjunction with the booster capacitor a high current at the beginning of the activation process, thus ensuring that the solenoid armature and the valve needle are raised quickly. In the pickup phase, the valve needle then achieves the maximum opening lift (Figure 13c). When the fuel injector is open (maximum valve-needle lift), a lower control current (holding current) is sufficient to keep the valve open. With a constant injector-needle lift, the injected fuel quantity is proportional to the injection duration (Figure 13d).

The CVO process (Controlled Valve Operation) has been specially developed to reduce the quantity scatter from valve to valve. The opening and closing times of the fuel injector can be determined from the electrical variables with this mechatronic approach. The current actual open duration is formed from these measured variables; this is delivered to a controller, where it is compared with a setpoint open duration. In this way, it is possible particularly for very small quantities to achieve high metering precision and maintain this level of precision with a downstream adaptation over the entire service life.

During multiple fuel injection the high-pressure fuel injector is actuated more than once during an engine cycle, reducing the level to which the internal en-

gine walls are moistened with liquid fuel. For the fuel injector this means achieving short idle times and more numbers of load cycles to be withstood over the operating time compared with single injection.

Piezo high-pressure fuel injector

Area of application

The piezo high-pressure fuel injector has been used since 2005 in complex combustion processes which impose specific requirements on the quality of the fuel spray and its stability over the engine's service life. Moreover, it combines outstanding very-small-quantity metering precision with high maximum injected fuel quantities and thus makes possible high specific power output without having to make compromises at low-load operating points. An example of highly-developed combustion processes is the spray-guided stratified-lean combustion process, which is achieved exclusively with the piezo high-pressure fuel injector in production vehicles.

Design and function

The piezo high-pressure fuel injector is operated at 20 MPa (200 bar) fuel pressure and comprises the function elements of valve assembly, actuator module, compensation element (“coupler”), and housing and connecting parts (Figure 14).

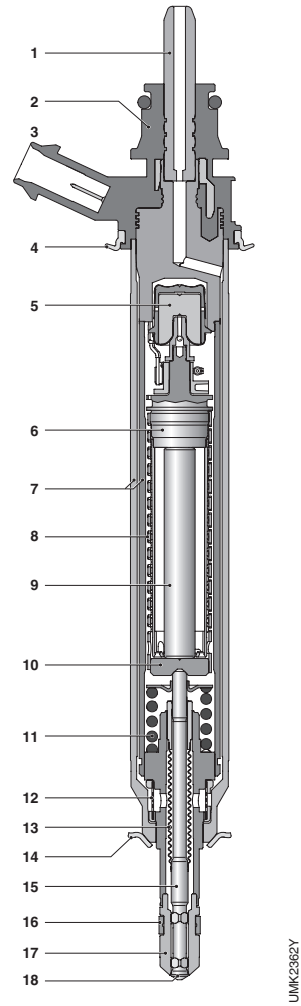
The valve assembly admits injected fuel into the combustion chamber and defines the shape of the spray. It essentially consists of the outward-opening valve needle which is pretensioned by a closing spring against a housing body. The nominal valve-needle lift (full lift) is 33 µm and is achieved after 180 µs. The closing operation also takes 180 µs from full lift.

Further important components of the valve assembly are a fine filter in the inlet to protect the nozzle against particulates and bellows to isolate the fuel from the dry actuator chamber.

Installed in the actuator chamber is the piezo actuator, which through its change of length under voltage opens and closes the valve needle and thus substantially determines fuel-quantity metering. It consists of over 400 active layers and is pretensioned by a tube spring. Aside from the actuator and tube spring, the electrical

Figure 14: Function elements of piezo fuel injector (Bosch)

1 Hydraulic port, 2 Plastic coating,
3 Electrical connection, 4 Injector-duct seal,
5 Linear compensation element,
6 Actuator foot, 7 Housing,
8 Piezo-actuator pretensioning spring,
9 Piezo actuator, 10 Actuator head,
11 Needle closing spring, 12 Mesh filter,
13 Metal bellows, 14 Support element,
15 Valve needle, 16 Combustion-chamber seal,
17 Valve body, 18 Injection nozzle.



contacting and insulation of the actuator against the electrical components are the essential elements of the actuator module. The direct non-positive connection between actuator module and valve needle provides for very short activation times from 75 µs. In the event of such very short activations the valve needle does not achieve its full lift.

The coupler exerts in the switching chain (comprising the valve needle, the actuator module, and the coupler itself) a defined force which counteracts the closing-spring force. In the coupler two steel diaphragms hold a hydraulic fluid which moves a working piston by means of a slow movement between two cavities. This brings about a compensation of the different thermal expansions of the metallic and ceramic components with the result that the coupler force within the switching chain remains to a large extent constant over all the thermal operating conditions of the high-pressure fuel injector. In this way, a reproducible valve-needle lift and thus a reproducible injected fuel quantity in the different engine operating states are made possible. Fast, short changes of switching-chain length by activation of the piezo actuator are not compensated by the coupler in contrast to changes of switching-chain length as a result of slow thermal effects. It works in this respect like a high-pass filter.

The above-mentioned modules are mounted together with the housing components including the fuel supply. A coding process integrated in the production sequence determines the individual charge demand of each piezo high-pressure fuel injector in order to compensate production tolerances. Compensating the production tolerances improves the precision of the fuel-quantity metering. The charge demand is written in the form of a data matrix code to the fuel injector and read later into the engine ECU.

Spray formation and spray targeting

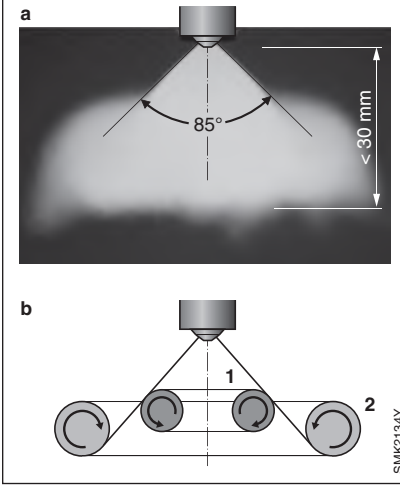
The spray of an outward-opening nozzle differs by virtue of the principle involved fundamentally from that of a multi-hole nozzle: The annular orifice created when the valve needle opens causes a tapered discharge of spray with droplet sizes in the region of 8 to 13 µm. A rotationally symmetric hollow taper with a nominal spray angle of 86 ° is created. Spatially permanently assigned boundary vortices, which can be used in the course of a combustion process, form both outside and inside along the taper surfaces (Figure 15). The spray shape is stable over a large combustion-chamber-pressure and charge-movement range such that even injections that shortly follow each other are possible in the compression stroke up to shortly before ignition (thus at high combustion-chamber pressures). The spray is, thanks to the design of the nozzle, very robust against disturbance, deposit, and aging influences. The is advantageous particularly for the stratified combustion process mentioned at the start, but also for example in catalyst-heating mode.

Electrical activation

The piezo high-pressure fuel injector is operated with up to 200 V. Activation is performed by an output stage integrated in the ECU, with the voltage generated by a

Figure 15: Tapered spray pattern of piezo fuel injector

Fuel pressure 200 bar,
chamber pressure 6 bar (lab measurement).
a) Realistic representation,
b) Schematic representation.
1 Inner boundary vortex,
2 Outer boundary vortex.



DC/DC converter (Figure 16). Activation is charge-controlled: The injector is opened by a specifically targeted charging operation (nominal value for full lift 0.69 mC) and closed by an analogous discharging operation. The activation duration (75 to 5000 µs) serves as the primary activation variable, the charge value to be achieved and the charging and discharging speeds are parameterizable. Varying the pressure to obtain an injected-fuel-quantity spread between minimum and maximum quantity is, due to the possible extremely short activation duration and on account of the high hydraulic cross-section at full lift, not necessary: Injected fuel quantities between 0.5 mg and 150 mg can be covered by a fuel injector per working cycle. The injection characteristic is linear right through to the smallest injected fuel quantities (Figure 17). Idle times between two opening operations right down to 50 µs can be achieved; this is considerably shorter than with a fuel injector with a solenoid-operated mechanism.

Finally, a number of adaptation functions which measure the fuel injector in the installed state are used in engine operation to guarantee very-small-quantity metering precision of the injector's service life. As well as model-based compensations of the temperature-dependent reactions of the actuator to an electric charge, it is also possible to compensate on a cylinder-specific basis very small quantity deviations over the injector's service life which if uncompensated would generate torque differences. The compensations act on the basis of the causative principle either on the valve-needle lift or on the activation duration.

Figure 16: Activation of piezo fuel injector
a) For full lift,
b) For partial lift.

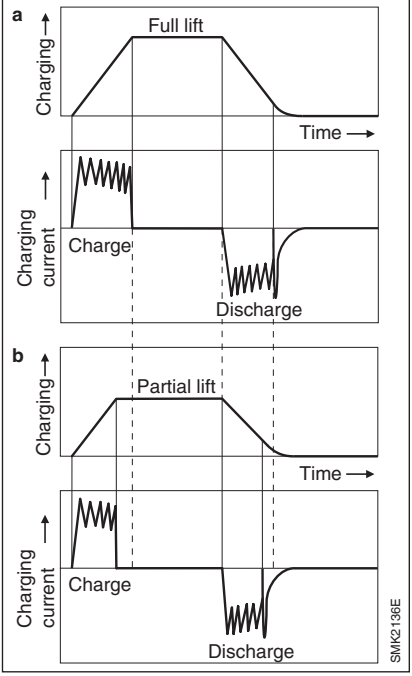
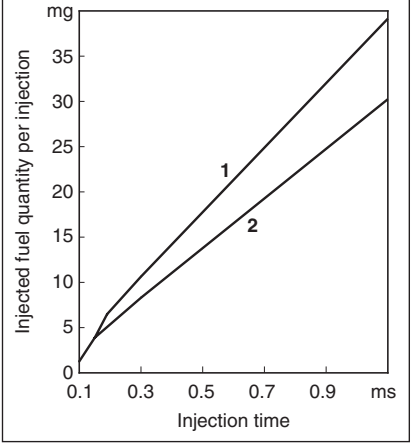


Figure 17: Injected fuel quantity as a function of the injection time
1 Full lift,
2 Partial lift.



Ignition

Basic principles

Function

On a spark-ignition (SI) engine, the combustion process is initiated by an externally supplied ignition. The ignition is responsible for igniting the compressed A/F mixture at the right time. This is done by producing an electric spark between the electrodes of a spark plug in the combustion chamber.

Consistent, reliable ignition under all conditions is essential to ensure fault-free engine operation. Misfiring leads to combustion misfires and damage or destruction of the catalytic converter, poor exhaust emission figures, higher consumption and lower engine performance.

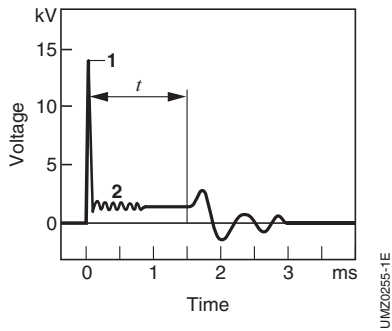
Ignition spark

An electric spark can only occur at the spark plug if the necessary ignition voltage is exceeded (Figure 1). The ignition voltage is dependent on the spark-plug electrode gap and the density of the A/F mixture at the moment of ignition. It can be up to several 10 kV. After flashover, the voltage at the spark plug drops to the firing voltage. The firing voltage depends on the length of the spark plasma (electrode gap and excursion by the A/F mixture flow) and can rise from several 100 V to a few kV.

During the ignition-spark combustion time (spark duration), the ignition-system energy is converted into the ignition spark. After the spark has broken away, the voltage is damped and drops to zero.

Figure 1: Spark-plug voltage characteristic with static or semi-static A/F mixture

1 Ignition voltage,
2 Firing voltage,
t Spark duration.



Mixture ignition and ignition energy

The electrical spark between the spark-plug electrodes generates a high-temperature plasma. If mixture conditions at the spark plug are suitable and sufficient energy has been supplied by the ignition system, the resulting arc develops into a flame front that propagates independently.

The ignition must guarantee this process under all engine operating conditions. Under ideal conditions (e.g. in a simulation of a combustion chamber, a “combustion bomb”), provided that the A/F mixture is stationary, homogeneous and stoichiometric, an energy of approx. 0.2 mJ is required to ignite the mixture by means of electric spark for each individual ignition process. In real engine operation, however, much higher energy levels are required. Some of the spark energy is converted at flashover, and the rest in the spark combustion phase.

Larger electrode gaps generate a larger arc, but require higher ignition voltages. Lean A/F mixtures or turbocharged engines have a higher ignition-voltage demand. At a given level of energy, the spark duration shortens as the ignition voltage increases. A longer spark duration generally stabilizes the combustion; lack of mixture homogeneity at the moment of ignition in the area of the spark plug can be compensated for by means of a longer spark duration. A/F-mixture turbulences such as those which occur in the stratified-charge mode with gasoline direct injection can divert the ignition spark to such an extent that it extinguishes. Follow-up sparks are then required to ignite the A/F mixture again.

The need for higher ignition voltages, longer spark durations, and the provision

of follow-up sparks have resulted in the design of ignition systems with higher ignition energy. If not enough ignition energy is produced, inflammation will not occur, resulting in engine misfire. The system must therefore deliver enough ignition energy to ensure reliable ignition of the A/F mixture under all operating conditions.

Efficient fuel atomization and good access of the A/F mixture to the ignition sparks enhance ignitability, extend spark duration and spark length, and lengthen the electrode gap. The dimensions of the spark plug determine the position (see Spark position, spark plug) and length of the spark; spark duration depends on the type and design of the ignition system, as well as on the instantaneous ignition conditions in the combustion chamber. Dependent on the engine requirements (intake-manifold injection, gasoline direct injection or exhaust-gas turbocharging), the spark energy of ignition systems lies in a range of approx. 30 to 100 mJ.

Moment of ignition

The start of combustion in a spark-ignition (SI) engine can be controlled by selecting the moment of ignition. The moment of ignition always refers to the top dead center of the spark-ignition (SI) engine power cycle. The earliest possible moment of ignition is determined by the knock limit, and the latest possible moment of ignition by the combustion limit or the maximum permissible exhaust-gas temperature. The moment of ignition affects the torque output, exhaust-gas emissions and fuel consumption.

Basic ignition point

Maximum combustion, and thus maximum combustion pressure, should occur shortly after top dead center (Figure 2) to deliver maximum engine torque. Since a certain amount of time passes from the moment of ignition to complete combustion of the A/F mixture, ignition must occur before top dead center. The moment of ignition must be advanced as rotational speed increases and charge decreases.

Likewise, the moment of ignition must be advanced in the case of lean A/F mixtures ($\lambda > 1$) because the flame front propagates more slowly. Recirculated or retained exhaust gas (external or internal exhaust-gas recirculation) has a compa-

Figure 2: Combustion-chamber pressure characteristic for various moments of ignition

1 Ignition (Z_a) at the correct point in time,
2 Ignition (Z_b) too advanced,
3 Ignition (Z_c) too retarded.

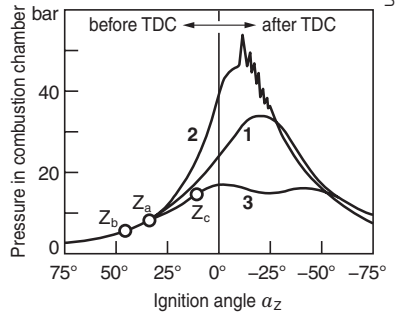
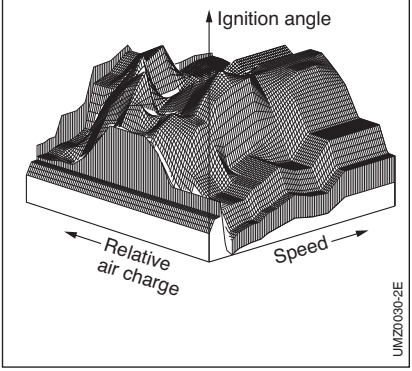


Figure 3: Ignition map as a function of engine speed and relative air charge



rable effect. Ignition-timing adjustment, therefore, essentially depends on rotational speed, charge, and the air/fuel ratio (excess-air factor λ) or in today's customary stoichiometric homogeneous engines the amount of residual exhaust gas in the cylinder. The moments of ignition are determined on the engine test bench and, in the case of electronic engine management systems, are stored in program maps (Figure 3).

Ignition-timing corrections and operating-point-dependent moments of ignition

Electronic engine-management systems can take other effects on the moment of ignition into consideration in addition to rotational speed and charge. The basic moment of ignition can either be modified by means of additive corrections or replaced for certain operating points or ranges by special ignition angles or ignition maps. Examples of ignition-timing corrections are knock control, the correction angle for the gasoline direct injection homogeneous-lean operation, and warm-up. Examples of special ignition angles or ignition maps are gasoline direct injection stratified-charge operation, and starting operation. Final implementation depends on the prevailing application.

Exhaust gas and fuel consumption

The moment of ignition has a considerable impact on exhaust gas because it can be used to control the various untreated exhaust-gas constituents directly. However, the various optimization criteria, such as exhaust gas, fuel economy, drivability, etc., may not always be compatible, so it is not always possible to derive the ideal moment of ignition from them.

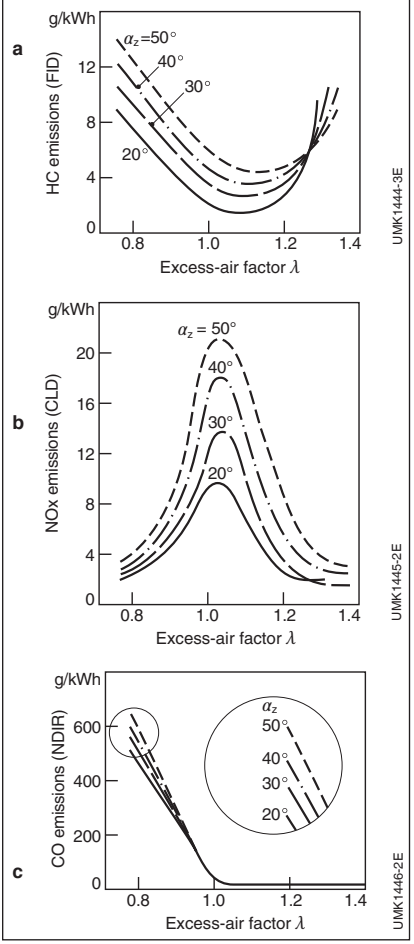
Shifts in the moment of ignition induce mutually inverse response patterns in fuel consumption and exhaust-gas emissions (Figure 4 and Figure 5). Whereas more spark advance increases torque and therefore power and reduces fuel consumption, it also raises HC and, in particular, nitrogen-oxide emissions. Excessive spark advance can cause engine knock that may damage the engine. Retarded ignition results in higher exhaust-gas temperatures, which can also harm the engine.

Electronic engine-management systems featuring programmed ignition curves are designed to adapt the moment of ignition in response to variations in factors such as rotational speed, load, temperature,

Figure 4: Influence of excess-air factor λ and moment of ignition α_z on pollutant emissions

a) Hydrocarbon emissions (HC),
b) Nitrogen-oxide emissions (NO_x),
c) Carbon-monoxide emissions (CO).

Flame Ionization Detector (FID)
Chemiluminescence Detector (CLD)
Non-Dispersive Infrared Detector (NDIR)



etc. They can thus be employed to achieve the optimum compromise between these mutually antagonistic objectives.

Knock control

Basic principles

Electronic control of the moment of ignition offers the possibility of accurate control of the ignition angle as a function of rotational speed, load, temperature, etc. Nevertheless, if there is no knock control, there must still be some means to define a clear safety margin to the knock limit.

This margin is necessary to ensure that, even in the most knock-sensitive case with regard to engine tolerances, engine aging, environmental conditions, and fuel quality, no cylinder can reach or exceed the knock limit. The resulting engine design leads to lower compression, retarded moments of ignition, and thus worsening fuel consumption and torque.

These disadvantages can be avoided through the use of knock control. Experience shows that knock control increases engine compression. This results in lower fuel consumption and higher torque. Now, however, the pilot control ignition angle no longer has to be determined for the conditions most sensitive to knocking but rather for the conditions least sensitive to knocking (e.g. compression of the engine at lowest tolerance limit, best possible fuel quality, cylinder least sensitive to knocking). Each individual engine cylinder can now be operated throughout its service life in virtually all operating ranges at its knock limit, and thus at optimum efficiency. For this type of ignition angle adjustment, a reliable method of knock detection is essential. It should detect knock for each cylinder throughout the engine's operating range starting from a specified knock intensity.

Knock-control system

A knock-control system consists of knock sensor, signal evaluation, knock detection, and ignition-angle control system with adaptation facility (Figure 6).

Figure 5: Influence of excess-air factor λ and moment of ignition α_z on fuel consumption and torque
a) Torque,
b) Specific fuel consumption.

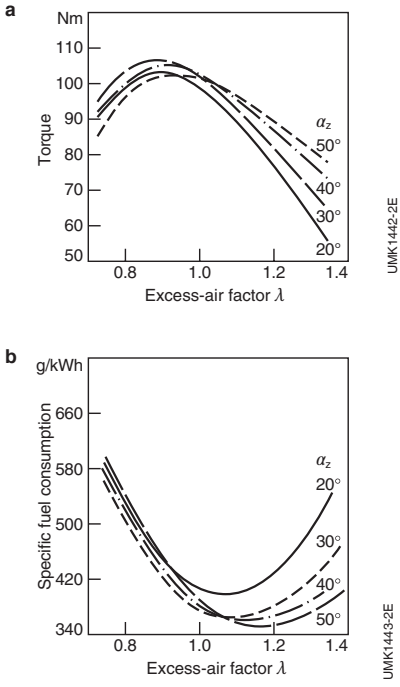
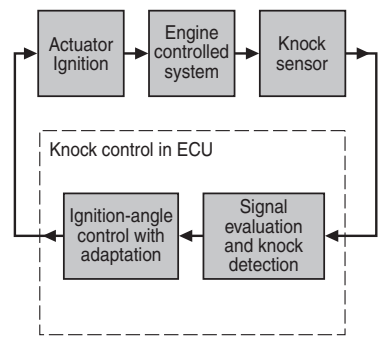


Figure 6: Schematic of a knock-control system



Knock sensor
A typical symptom of combustion knock is high-frequency vibrations which are superimposed on the high-pressure curve in the combustion chamber (Figure 2). These vibrations are best detected directly in the combustion chamber by means of pressure sensors. As fitting these pressure sensors in the cylinder head for each cylinder is still relatively costly, these vibrations are usually picked up using knock sensors fitted to the exterior of the engine. These piezo-electric acceleration sensors (Figure 7) pick up the characteristic vibrations of knocking combustion and convert them into electrical signals.

There are two types of knock sensor. A wide-band sensor, with a typical frequency band of 5 to 20 kHz, and a resonance sensor, which preferably transmits only one knock-signal resonant frequency. When combined with the flexible signal-evaluation system in the control unit, it is possible to evaluate different or several resonant frequencies from one wide-band knock sensor. This improves knock detection performance, which is why the wide-band knock sensor is increasingly replacing the resonance sensor.

To ensure sufficient knock detection in all cylinders and across all operating ranges, the number and location of the required knock sensors must be carefully determined for each engine type. Four-cylinder in-line engines are usually fitted with one or two knock sensors, while 5- and 6-cylinder engines are fitted with two, and 8- and 12-cylinder engines with four knock sensors.

Signal evaluation
For the duration of a timing range in which knock can occur, a special signal evaluation circuit in the control unit evaluates from the wide-band signal the frequency band(s) with the best knock information and generates a representative variable for each combustion process. This very flexible signal evaluation of the wide-band sensor enables high detection quality. When using a resonance knock sensor that transmits just one resonant frequency for analysis of all cylinders assigned to it across the entire engine map, knock detection is normally no longer possible at higher engine speeds.

Knock detection
The variable produced by the signal-evaluation circuit is classified in a knock-de-

Figure 7: Knock sensor
1 Seismic mass, 2 Sealing compound,
3 Piezoceramic, 4 Connection,
5 Electrical connection.

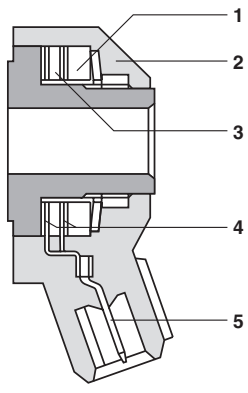
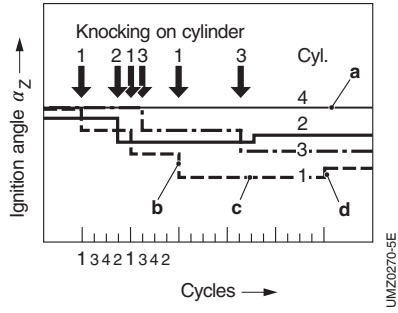


Figure 8: Knock control
Control algorithm for ignition adjustment on a 4-cylinder engine.
$K_{1\ to\ 3}$ Occurrence of knocking at cylinders 1 to 3, at cylinder 4 no knocking.
a Map ignition angle,
b Ignition retard step with knock,
c Ignition advance waiting time,
d Ignition advance step.



tection algorithm as “knock” or “no knock” for each cylinder and for each combustion process. This is done by comparing the variable for the current combustion process with a variable which represents combustion without knock.

Ignition-angle control system with adaptation facility

If combustion knock is detected in a cylinder, the moment of ignition for that cylinder is retarded (Figure 8). When knock stops, the moment of ignition is advanced again in stages up to the precontrol value. The knock-detection and knock-control algorithms are matched in such a way as to eliminate any knock that is audible and damaging to the engine, even though each cylinder is operated at knock limit within the optimum efficiency range.

Real engine operation produces different knock limits, and thus different moments of ignition for individual cylinders. In order to adapt precontrol values for the moment of ignition to a particular knock limit, the ignition retard values are stored for each cylinder dependent on the operating point. They are stored in non-volatile program maps in the permanently powered RAM for load and engine speed. In this way, the engine can be operated at optimum efficiency for each operating point and without audible combustion knocks, even if there are rapid load and engine-speed changes.

This adaptation even enables the use of fuels with lower antiknock properties (e.g. premium instead of Super Plus gasoline).

Phenomenon of pre-ignition

Basic principles

There is currently a clear trend in the development of modern gasoline engines towards downsizing (displacement reduction at same engine performance) and downspeeding (speed reduction by means of longer transmission ratio) in combination with direct injection and supercharging. Supercharging allows for a reduction in displacement without lowering performance level. The engine can therefore be operated in part load at higher loads with higher part-load efficiency, and fuel consumption can be reduced. The charge-air pressure increase to improve efficiency is however limited by the phenomenon of pre-ignition.

A few years ago the terms “extreme knocker” or “superknocker” were used, but they described the symptom and not the cause of the phenomenon of pre-ignition. However, the term “pre-ignition” has become the established term for some time now. In the name there is therefore also a clear distinction from normal combustion knock.

Pre-ignition

Pre-ignitions are uncontrolled auto-ignitions of the air-fuel mixture, which occur before the ignition that is triggered by the spark plug. Compared to normal combustion, this premature initiation of combustion leads to a considerable increase in pressure and temperature, which then leads to severe combustion knock in the subsequent combustion process. In doing so, individual events can pre-damage the engine.

Detection of pre-ignitions

Reliably detecting pre-ignitions is absolutely necessary in order to prevent engine damage. Knocking events as a result of pre-ignitions can be clearly evaluated and detected based on knock sensors by means of position and frequency ranges. This process allows for a high level of detection quality with a clear distinction from normal combustion knock.

Measures to prevent pre-ignitions
If pre-ignitions are detected, countermeasures are necessary in order to prevent further pre-ignitions. However, since pre-ignition is an uncontrolled auto-ignition, there is no direct manipulated variable – e.g. the ignition angle at combustion knock – that can be used to reliably prevent further pre-ignitions. For this reason, the pre-ignition functionality involves initiating multiple countermeasures after detected pre-ignitions in order to quickly reduce the temperature level in the combustion chamber and thus prevent further pre-ignitions. The various measures, such as mixture enrichment or charge lowering, can be activated in the combination that is ideal for the respective engine type.

Pre-ignitions are heavily dependent on fuel and on oil emissions. This functionality is therefore absolutely necessary to protect the engine, especially for turbocharged gasoline engines. This enables vehicle manufacturers to develop efficiency-optimized, supercharged gasoline engines and to market them worldwide despite different fuel and oil grades.

Risk of accidents
All electronic ignition systems are high-voltage systems. To avoid any risks, always switch off the ignition, or disconnect the power supply when working on the ignition system. Such work includes, for example:
– Replacing parts such as spark plugs, ignition coils, or ignition transformers, high-tension ignition cables, etc.
– Connecting engine analyzers, such as stroboscopic lamps, dwell angle/rotational speed testers, ignition oscilloscopes, etc.

When checking the ignition system, remember that dangerously high levels of voltage are present within the system whenever the ignition is on. All tests and inspections should therefore only be carried out by qualified professional personnel.

Ignition systems

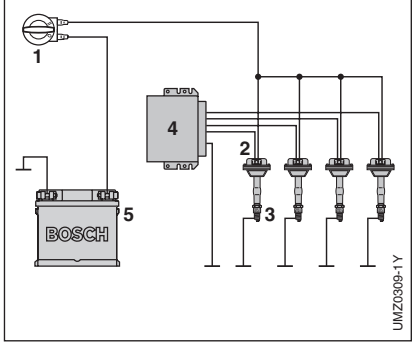
On modern vehicles, the ignition systems are almost always incorporated as subsystems of the engine-management system. Autonomous ignition systems are now only used for special applications (e.g. small engines). In the case of ignition systems, coil ignition (inductive ignition) with a separate ignition circuit per cylinder (static high-voltage distribution with single-spark coils, Figure 9) has come to the fore.

Alongside this, but to a much lesser extent, high-voltage capacitor ignitions (capacitive ignition) or other special designs such as magnetos are used for small engines. The next section will focus on coil ignition alone.

Coil ignition (inductive ignition)

Principle of coil ignition
The ignition circuit of a coil-ignition system (Figure 10) consists of:
– An ignition coil with a primary and a secondary winding
– An ignition output stage to control the current by means of the primary winding, predominantly integrated as an IGBT (Insulated Gate Bipolar Transistor) in the engine control unit or in the ignition coil
– A spark plug connected to the high-voltage connection point of the secondary winding

Figure 9: Ignition system with single spark coils
1 Ignition lock, 2 Ignition coil, 3 Spark plug, 4 Control unit, 5 Battery.



Before the desired moment of ignition, the ignition output stage switches a current from the vehicle electrical system through to the primary winding of the ignition coil. While the primary current circuit is closed (dwell period), a magnetic field builds up in the primary winding.

At the moment of ignition, the current through the primary winding is interrupted again, and the magnetic-field energy is discharged, mainly via the magnetic-coupled secondary winding (induction). In the process, a high voltage is produced in the secondary winding. Flashover occurs if the ignition voltage supply of the ignition system exceeds the ignition-voltage demand of the spark plug. After flashover, the remaining energy is converted at the spark plug while the spark is present.

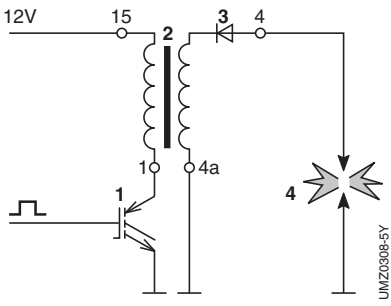
Functions of an ignition system with coil ignition

Moment-of-ignition determination

The current moment of ignition in each case is determined from program maps dependent upon the operating point and output.

Figure 10: Structure of an ignition circuit with single spark coils

1 Ignition output stage,
2 Ignition coil with primary and secondary windings,
3 Activation arc diode (Activation Spark Suppression),
4 Spark plug.
15, 1, 4, 4a Terminal designations,
⎍⎍ Triggering signal.



Dwell-period determination

The required ignition energy is made available at the moment of ignition. The amount of ignition energy is dependent upon the amount of primary current at the moment of ignition (cutoff current) and the inductance of the primary winding. The amount of cutoff current is mainly dependent upon the operating time (dwell period) and upon the battery voltage at the ignition coil. The dwell periods required to achieve the desired cutoff current are contained in characteristic curves or program maps as a function of the steady-state voltage. The change in dwell period with temperature can also be compensated for.

Ignition release

The ignition release ensures that the ignition spark occurs at the right cylinder at the right time and with the required level of ignition energy. On electronic-controlled systems, a trigger wheel with a fixed-angle reference mark (typically 60 – 2 teeth, two teeth missing at the reference mark) on the crankshaft is usually scanned (sensor system with hall or inductive sensor). From this, the control unit can calculate the crankshaft angle and the momentary rotational speed. The ignition coil can be switched on and off at any required crankshaft angle. An additional phase signal from the camshaft is required for the unambiguous identification of the cylinder.

For each combustion, the control unit uses the desired moment of ignition, the required dwell period and the current engine speed to calculate the switch-on time and switches on the output stage. The moment of ignition, or the switchoff point for the output stage, can be triggered either when the dwell period expires or when the desired angle is reached.

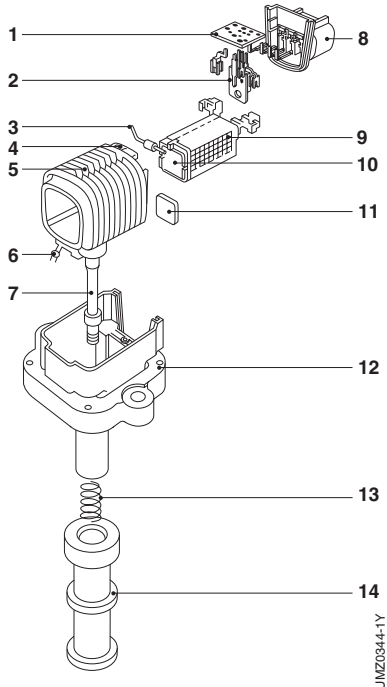
Ignition coil

Function

The ignition coil is principally an energy-charged high-voltage source similar in structure to a transformer. Energy is supplied by the vehicle electrical system during the dwell period or charging time. At the moment of ignition, which at the same time is the end of the charging time, the energy is then transferred with the required high voltage and sparking energy to the spark plug (inductive ignition system).

Figure 1: Design of the compact ignition coil

1 Printed-circuit board (optional),
2 Ignition output stage (optional),
3 Activation arc diode (optional),
4 Secondary-winding housing,
5 Secondary winding, 6 Contact plate,
7 High-voltage pin
(connecting element to contact spring),
8 Connector, 9 Primary winding,
10 I-core, 11 Permanent magnet, 12 O-core,
13 Contact spring (spark-plug contacting),
14 Silicone jacket (high-voltage insulation).



Design

The ignition coil (Figure 1) comprises two windings that are magnetically linked by an iron core (I-core and O-core). This iron core may contain a permanent magnet for energy optimization. Compared with the secondary winding, the primary winding has significantly fewer coils. The turns ratio $\ddot{u}$ is in the range of $\ddot{u}$ = 80 to 150.

The windings must have good electrical insulation to prevent electrical discharge and flashovers either to the inside or to the outside. For this purpose, the windings are usually cast in epoxy resin in the ignition coil housing.

As a rule, the iron core consists of stacked, ferromagnetic steel-plate fins in order to minimize above all eddy-current losses.

Alternatively, the ignition output stage can be incorporated into the ignition coil instead of the engine control unit. Interference-suppression elements may also be incorporated into the ignition coil together with the activation arc diode (activation spark suppression). An interference-suppression resistor is commonly used on the high-voltage output to the spark plug.

Operation

The ignition output stage switches the primary current in the ignition coil. The current rises with a delay, in accordance with the inductance. Energy is stored in the magnetic field created in the process. The operating time (charging time) is calculated in such a way that a specific cutoff current, and thus a specific level of stored energy, will be reached by the end of the operating time.

The current is cut off by the ignition output stage, which causes a swift change in the magnetic flux in the ignition coil's iron core. This change in flux results in a voltage induction in the secondary winding. The design of the secondary winding, the geometric arrangement of the secondary winding to the iron core and the primary winding and the materials used in the insulation system and the iron core give rise to inductive and capacitive properties that lead to a voltage supply in excess of 30,000 V on the secondary side.

When the ignition-coil voltage supply equals the ignition-voltage demand at the spark-plug electrodes, the voltage drops to

a spark voltage of about 1,000 V. A spark current then flows and decreases as spark duration increases, until the spark is finally extinguished.

As the current changes over time when the primary current is cut off, an induction voltage occurs at the ignition coil output, similar to the start of charging time. However, this voltage is much smaller than the voltage at the moment of ignition and has an inverse polarity to that voltage. To prevent this "transient voltage" causing unwanted ignition, it is usually suppressed by means of a high-voltage diode in the secondary circuit (activation arc diode, activation spark suppression).

The design of an ignition coil can determine its electrical characteristics. In this respect, the requirements in terms of installation space (e.g. geometry of cylinder head and cylinder-head cover, position of fuel injector and intake manifold) and the two specified interfaces, the ECU with output stage (e.g. cutoff current) and the spark plug (e.g. ignition voltage, spark data), are decisive.

Design variations

There are several types of ignition coil that are distinguishable by their various features.

Individual coils and modules

In addition to the individual coils, which normally sit directly on the spark plug, several ignition coils can be grouped together in a single housing as a module. They are then mounted directly on the spark plugs, or they may be located a short distance away, in which case the high voltage must be supplied via appropriate cables.

Single-spark and dual-spark ignition coils

In addition to ignition coils with just one high-voltage output (single-spark ignition coil), there are also coils that use both ends of the secondary winding as an output (dual-spark ignition coil). The electric circuit must always be closed by both spark gaps to prevent discharge on the secondary side. One feasible application for this design is dual ignition, i.e. two spark plugs per cylinder supplied by one ignition coil. Another application involves splitting up the two high-voltage outputs between two spark plugs on different cylinders. In this case, one of the two spark plugs will always be in the ignition stroke. As a result, the voltage and energy requirements of the "passive spark" (backup spark) are reduced substantially. Above all, this variant offers cost advantages; however, it must be matched to the overall system to prevent damage from unwanted ignition events caused by backup sparks.

Compact coil and pencil coil

Ignition coils are also distinguishable by their basic design. For example, there is the conventional compact coil, that features an equal-sided coil body and an O-I core or C-I core magnetic circuit. The coil body sits in the engine above the spark-plug well.

Another type is the pencil coil whose coil body projects into the spark-plug well. Here, too, the windings are located on an I or pencil core, with a plate (yoke plate) arranged concentrically around the windings serving as a magnetic yoke.

Ignition coils with ignition output stage

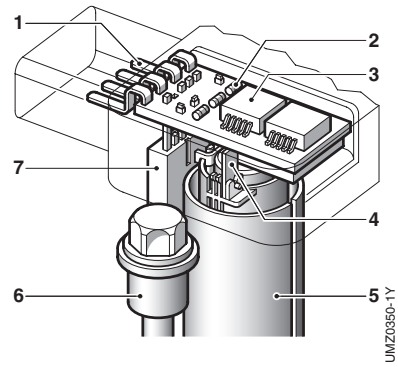
Ignition coils are available with or without the ignition output stage. The reason for integrating the output stage is to relieve the strain on the engine ECU. Additionally, further electronic circuitry is used in the ignition coil in order to meet the existing requirements (Figure 2).

Requirements
The principal requirements of modern ignition systems are derived indirectly from necessary emission and fuel reductions. Requirements of ignition coils are derived from relevant engine solutions, such as high-pressure supercharging and lean-burn and stratified-charge operation (spray-guided direct injection) in combination with increased exhaust-gas recirculation rates (EGR).

It is necessary to represent an increased ignition-voltage and energy demand with increased temperature requirements. This is effected among others by means of
– High-energy coils with high voltage supply (in excess of 40,000 V)
– Multispark ignition (MSI), i.e. spark-sequence control with integrated electronics (ASIC), controlled by way of primary- and secondary-current evaluation
– Diagnostic functions (e.g. charge-time monitoring, ionic-current measurement for combustion diagnosis)
– Protective functions (e.g. overtemperature cutoff, cutoff-current regulation)

Furthermore, motor vehicles are subject to increased demands with regard to electromagnetic compatibility (EMC). Precisely because of the increased ignition-voltage demand, the increased ignition frequencies necessitated by multispark ignition and the increased cutoff currents, it is now necessary to reduce the interference emitted by the ignition system so as not to disrupt the functioning of other automotive components (ECUs, microcontrollers, sensors, actuators).

Figure 2: Installation of ignition output stage in a pencil ignition-coil housing
1 Connector,
2 SMD components (Surface Mounted Device),
3 Electronic circuitry for ignition functions,
4 Primary-winding contacts,
5 Yoke plate of pencil ignition-coil transformer,
6 Attachment lug,
7 Ignition output stage.



The integration of increasingly more electronic circuitry in ignition coils gives rise to more stringent demands with regard to the interference immunity of these components. In the interests of avoiding impaired functioning and malfunctions, the electronic circuity must be immune to the emissions of the ignition system itself and to the interference emitted by other vehicle components.

Spark plug

Function

The spark plug transfers the ignition-coil energy into the combustion chamber in a spark-ignition engine. The high voltage applied creates an electric spark between the spark-plug electrodes, which ignites the compressed A/F mixture and thereby initiates combustions.

In combined operation with the other engine components, e.g. the ignition and the mixture-formation systems, the spark plug determines the function of the spark-ignition engine to a decisive extent. It must provide for reliable cold starting, guarantee misfire-free operation over its entire service life (at least 30,000 km), and not exceed the maximum permissible temperature even during extended operation at critical operating points. Furthermore, the spark plug must always reliably insulate the high voltage against the cylinder head and seal the combustion chamber towards the outside.

Requirements

The spark plug must satisfy a variety of challenging performance demands. It is exposed to the varying periodic processes within the combustion chamber, as well as external climatic conditions.

When spark plugs are used with electronic ignition systems, ignition voltages of up to 40 kV can occur; these high voltages must not cause ceramic or insulator-head flashovers. The deposited residues from the combustion process such as soot, carbon and ash from fuel and oil additives are electrically conductive under certain thermal conditions. Nevertheless, no flashovers via the insulator are permitted to occur. The electrical resistance of the insulator must be sufficiently high up to 1000 °C and is only permitted to drop slightly over the spark plug's service life.

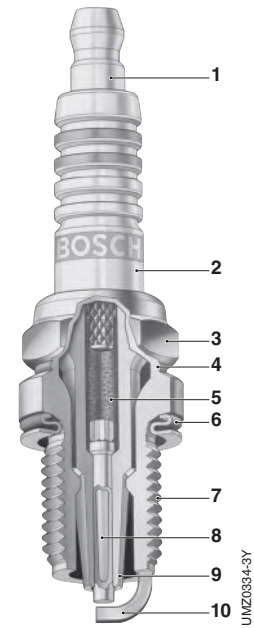
Mechanically, the spark plug is subjected to the pressures (up to 150 bar) occurring periodically in the combustion chamber; however, gas tightness may not be impaired. In addition, the spark-plug electrode materials exhibit extreme resistance to thermal loads and continuous vibratory stress. The spark-plug shell must be able to absorb the forces occurring during assembly without any lasting deformation.

At the same time, the section of the spark plug that protrudes into the combustion chamber is exposed to high-temperature chemical processes, making resistance to aggressive combustion deposits essential.

Above all, the spark-plug insulator and the electrodes must be capable of withstanding high thermal loads as they are exposed in rapid succession to high temperatures from the hot combustion gases and low temperatures of the cold air/fuel mixture (thermal shock). The electrodes and the insulator at the cylinder head must have good heat dissipation properties – essential for reliable spark-plug perfor-

Figure 1: Spark-plug design

1 Terminal stud
(here with SAE terminal),
2 Insulator head,
3 Nickel-plated steel shell,
4 Heat-shrinkage zone,
5 Electrically conductive glass seal,
6 Captive gasket (seal seat),
7 Thread,
8 Compound center electrode (Ni/Cu),
9 Insulator nose,
10 Ground electrode
(here Ni/Cu compound electrode).



mance. The connection end of the spark plug should heat up as little as possible.

As well as these requirements, the spark plug must also be adapted to the geometric specifications of the engine design (e.g. spark-plug position in the cylinder head).

In view of these requirements – resulting from the great variety of engines – a wide range of spark plugs is required.

Design

The spark plug consists of at least two electrodes (center and ground electrodes), between which the ignition spark is generated. The high voltage is transferred to the center electrode via the terminal stud by an electrically conductive glass seal inside the insulator. The insulator prevents a short-circuit to the spark-plug shell, in which the plug (insulator with operationally reliable mounted center electrode and terminal stud) is mounted (Figure 1).

Terminal stud and high-voltage terminal

The steel terminal stud has at the end projecting out of the insulator a thread (M4) into which the spark-plug connector of the ignition cable engages. For the standardized connector either a terminal nut (SAE) is screwed onto the terminal-stud thread or the stud is already provided during manufacture with a solid standardized terminal (e.g. cup). Spark plugs with metal shields are available for watertight systems and for maximum interference suppression.

Conductive glass seal

This electrically conductive glass seal acts as a mechanical support for the components, while providing a gas seal against the high combustion pressure. It can also incorporate resistor elements for interference suppression and burn-off.

Insulator

The insulator consists of a special ceramic. Its function is to insulate the center electrode and the terminal stud electrically against the spark-plug shell. The requirements of good thermal conductivity with simultaneously high insulation resistance are in stark contrast to the properties of most insulating materials. The material used by Bosch consists of aluminum oxide (Al_2O_3), to which other substances are added in small quantities.

At the connection end, the insulator has a lead-free glaze to repel moisture and dirt. This avoid insulator flashovers to a great extent.

Spark-plug shell

The shell is made of steel. The lower part of the shell has a thread so that the spark plug can be secured in the cylinder head and exchanged after a specified change interval. Depending on the spark-plug concept, up to four ground electrodes are welded onto the end face of the shell.

To protect the shell against corrosion, its surface is nickel-plated so as to prevent the thread from seizing in the aluminum cylinder heads. The upper part of the shell features a hexagon or bi-hexagon head for application of a wrench. The bi-hexagon head requires less space in the cylinder head while the insulator-head geometry remains unchanged, allowing the engine design engineer greater freedom in designing the cooling ducts.

The upper part of the shell is flanged after the plug is fitted and locates the latter in its position. The subsequent shrinking process by inductive heating under high pressure creates the gastight connection between the insulator and the shell, and guarantees good heat conduction.

Seal seat
Depending on the engine design, the spark plug has a flat or conical seal seat, which provides the seal against the cylinder head. In the case of the flat seal seat, a "captive" sealing ring is used as the sealing element. It has a special shape and provides a permanently elastic seal when the spark plug is fitted. In the case of the conical seal seat, a conical shell surface provides a direct seal on a corresponding surface of the cylinder head without the use of a sealing ring.

Electrodes
The electrodes subjected to high thermal load are primarily made of a nickel-based multi-alloy. Through the addition by alloying of manganese and silicon the chemical resistance of nickel is improved, above all against the highly aggressive sulfur dioxide (SO_2, sulfur is an ingredient of lube oil and fuel). Additives made of aluminum and yttrium additionally increase the resistance to scaling and oxidation.

Compound electrodes with a jacket material made of a nickel alloy and a copper core are used to improve heat dissipation and consequently wear performance.

Center electrode
The center electrode is anchored with its head in the conductive glass seal. For longlife spark plugs it is sensible to use corrosion- and oxidation-resistant materials like platinum (Pt) and platinum alloys, which also exhibit high burn-off resistance. The center electrode then accommodates a precious-metal pin, which is permanently joined to the base electrode by laser welding.

Ground electrodes
The ground electrodes are welded to the shell and have predominantly a rectangular cross-section. Depending on the arrangement, a distinction is made between front and side electrodes (Figure 3) and special applications (e.g. spark plug without distinct ground electrode for racing engines). As well as thermal conductivity, the length, the profile cross-section and the number of ground electrode determine their temperature and hence their wear performance.

Heat range
Operating temperature of the spark plug
When the engine is operating, the spark plug is heated by combustion heat. Some of the heat absorbed by the spark plug is diverted to the fresh A/F mixture. Most of the heat is transmitted to the spark-plug shell via the center electrode and the insulator, and is diverted to the cylinder head. The operating temperature of the spark plug represents a balance between heat absorption from the engine and heat dissipation to the cylinder head.

The aim is for the insulator nose to reach a self-cleaning temperature of approx. 500 °C even at low engine performance. If the temperature drops below this level, there is the danger that soot and oil residue from incomplete combustion will settle on the cold areas of the spark plugs (particularly when the engine is not at normal operating temperature, at low outside temperatures and during repeated starts) (Figure 2, curve 3). This can create a conductive connection (shunt) between the center electrode and the spark-plug shell. This will cause ignition energy to leak away in the form of short-circuit current (risk of misfiring).

At higher temperatures, the residue containing carbons burn on the insulator nose; the spark plug thus "cleans" itself (Figure 2, curve 2).

An upper temperature limit of approx. 900 °C should be observed since, in this range, spark-plug electrode wear increases drastically (due to oxidation and hot-gas corrosion). If this limit is exceeded by a significant extent, it increases the risk of auto-ignition (ignition of the A/F mixture on hot surfaces) (Figure 2, curve 1). Auto-ignition subjects the engine to extreme loads, and may result in engine destruction within a short period of time. The spark plug must therefore be adapted to the engine type in terms of its heat-absorbing property.

Heat-range code number
The identifying feature of a spark plug's thermal loading capacity is its heat range, which is defined by a heat-range code number and determined in comparison measurements with a reference standard source (calibration spark plug).

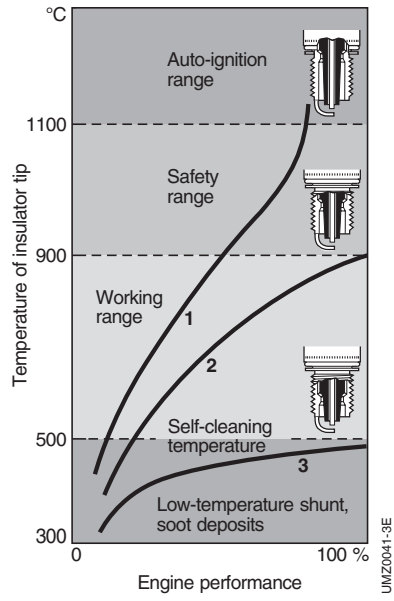
A low code number (e.g. 2 to 5) denotes a "cold spark plug" with low heat absorption through a short insulator nose. High heat-range code numbers (e.g. 7 to 10) denote "hot spark plugs" with high heat absorption through long insulator noses.

The heat range matching the respective engine is determined in the course of application measurements during development.

Ionic-current measurement procedure
The Bosch ionic-current measurement procedure employs the combustion process as a factor for determining the heat-range requirement. The ionizing effect of flames is used to assess how combustion develops over time; this is done by measuring conductivity in the spark gap.

Figure 2: Spark-plug temperature response
1 Spark plug with excessively high heat-range code number (hot plug),
2 Spark plug with suitable heat-range code number,
3 Spark plug with excessively low heat-range code number (cold plug).

The working range temperature should be from 500 °C to 900 °C for different engine performance ratings.



Characteristic changes in the combustion process due to increased thermal loading of the spark plugs can be detected using ionic current and used in the assessment of the auto-ignition process.

Thermal ignition
Ignitions of the air/fuel mixture which occur independently of the ignition spark and for the most part on a hot surface (e.g. on the excessively hot insulator-nose surface of a spark plug with excessively high heat range) are called auto-ignitions. These can be subdivided into two categories on the basis of their position in time relative to the moment of ignition.

Post-ignition
Post-ignitions occur after the electrical moment of ignition, but are not critical for practical engine operation since electrical ignition occurs at an increasingly earlier stage. To find out whether thermal ignitions are initiated by the spark plug, individual ignitions are cyclically suppressed during the ionic-current measurement. When a post-ignition occurs, the ionic current increases significantly only after the moment of ignition. But since a combustion is initiated, a pressure increase and hence a torque output are also to be observed.

Pre-ignitions
Pre-ignitions occur before the electrical moment of ignition and can, on account of their uncontrolled progression (see Irregular operating states), cause serious engine damage. Excessively early combustion initiation shifts not only the position of the pressure maximum to top dead center (TDC), but also the maximum combustion-chamber pressure to higher values. This increases thermal and mechanical load on the components in the combustion chamber.

Spark-plug choice
The aim of an adaptation is to choose a spark plug which can be operated without thermal ignitions before the moment of ignition (pre-ignition) and has a sufficient heat-range reserve. In this way, the variations in engine and spark-plug production are covered and it is also taken into consideration that engines can change in terms of

their thermal properties over the course of their operating life. Thus, for example, oil-ash deposits in the combustion chamber can increase the compression ratio, which in turn causes a higher temperature load on the spark plug.

The correct heat range for the engine is determined when no failures with sooted spark plugs occur in the concluding cold-start tests with this heat-range recommendation.

The use of materials with higher thermal conductivity for the center electrodes (silver or nickel alloys with copper core) provides – with the same heat-range code number – for a significant increase in the insulator-nose length, the upshot of which is that the spark plug is less prone to heavy sooting. This reduces the likelihood of combustion misses and misfiring – with their attendant massive increases in hydrocarbon emissions – and provides benefits in exhaust emissions and fuel consumption in part-throttle operation at low load factors.

Close cooperation between engine and spark-plug manufacturers is customary when it comes to choosing suitable spark plugs.

Electrode gap and ignition voltage

The electrode gap is the shortest distance between the center electrode and the ground electrode (Figure 3). It determines, among other things, the ignition-spark length and should on the one hand be as large as possible to enable the ignition spark to activate a large area of the air/fuel mixture. This gives rise through stable flame-core formation to a reliable ignition of the air/fuel mixture. On the other hand, a lower ignition voltage is needed to produce a spark with a smaller electrode gap. If the electrode gap is too small, however, only a small flame core will be created around the electrode. Since energy will be drawn from the flame core via the contact areas with the electrodes (quenching), the flame core will only be able to propagate very slowly. In extreme cases, so much energy may be drawn off that misfires will occur.

During flashover and operation at higher temperature the electrode material is subjected to such high stress that the electrodes wear and consequently the electrode gap gets bigger. As the electrode gap increases the ignition conditions are in fact improved, but the ignition-voltage demand increases. As the ignition coil's voltage supply is a fixed amount, the ignition-voltage reserve is reduced and the risk of misfiring increases.

The required ignition voltage is influenced not only by the electrode gap, the electrode shape, the temperature, and the electrode material, but also by parameters specific to the combustion chamber, such as mixture composition (λ value), flow velocity, turbulence, and density of the ignitable gas.

In modern engine concepts with high mixture density, which frequently feature high charge turbulence, electrode gaps must be carefully defined in order to guarantee reliable ignition, and thus misfire-free operation throughout the required service life.

Spark position

The position of the spark gap relative to the combustion chamber wall defines the spark position. On modern engines (and particularly with gasoline direct-injection engines), the spark position has a considerable influence on combustion. A perceptible improvement in ignition response is observed when the spark position projects deeper into the combustion chamber. Combustion can be characterized by smooth-running engine performance, which can be derived directly from engine speed fluctuations.

However, because the ground electrodes are longer, higher temperatures are achieved. This, in turn, has an effect on electrode wear and electrode durability. It is possible to achieve the required service life by implementing design measures (e.g. extending the spark-plug shell beyond the combustion chamber wall), or by using compound electrodes, or high-temperature-resistant materials.

Spark-plug concepts

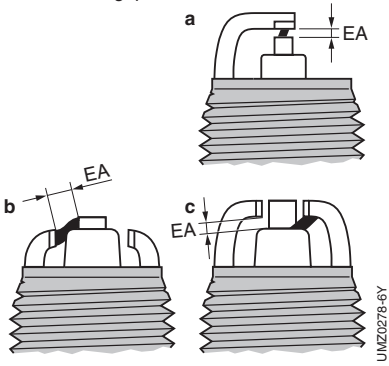
One or more ground electrodes can be advantageous, depending on the demands placed on the spark plug (wear, ignition response, etc.). The spark-plug type is determined by the relative location of the electrodes to each other and the position of the ground electrodes with respect to the insulator.

Air-gap spark concept
In the air-gap spark concept (Figure 3a) the ignition spark crosses between the center electrode and the ground electrode by the direct route and ignites the air/fuel mixture.

Surface-gap spark concept
The position of the ground electrode relative to the ceramic produces the surface-gap spark concept, in which the spark travels from the center electrode across the surface of the insulator nose before jumping across a gas-filled gap to the ground electrode (Figure 3b). The advantages here are the lower voltage demand with the same electrode gap, improved ignition properties,and the insulator-cleaning effect of the surface-gap spark for better repeat cold-starting performance.

Surface-air-gap spark concept
When both forms of discharged are made possible by the specific geometric arrangement of the ground electrodes, this is called the surface-air-gap spark concept (Figure 3c). Depending on operating conditions, the spark behaves either as an air-gap spark or a surface-gap spark with different ignition-voltage demand values.

Figure 3: Spark-plug concepts
a) Air-gap spark concept with front electrode,
b) Surface-gap spark concept with side electrodes,
c) Surface-air-gap spark concept with side electrodes.
EA Electrode gap.



In view of the trend towards increasing combustion-chamber pressures, air-gap spark concepts are to be preferred since these do not cause spark embedding in the insulator.

Simulation-based development of spark plugs
The Finite-Element Method (see FEM examples) is used for spark plugs to calculate temperature and electric fields and to solve structure-mechanical problems. Changes to the geometry and materials of spark plugs or even different physical boundary conditions and their effects can thus be determined in advance without complicated tests. This is the basis for a targeted production of test specimens, with which the calculation results are verified by way of example.

Spark-plug operating performance
Electrode wear
As spark plugs operate within an aggressive atmosphere, sometimes at extremely high temperatures, electrodes are subject to wear. This material erosion causes the electrode gap to increase noticeably the longer the spark plug is in service, therefore causing the ignition-voltage demand to increase. When this requirement can no longer be met by the supply from the ignition coil, misfiring will occur.

There are essentially two mechanisms that are responsible for electrode wear: spark erosion and corrosion in the combustion chamber. The flashover of electric sparks results in an increase in the temperature of the electrodes up to their melting temperature. The fused microscopically small surface areas react with the oxygen or the other constituents of the combustion gases. The consequence is material erosion.

Materials with a high thermal resistance (special-steel alloys of platinum or iridium) are used to minimize electrode wear. But material wear can also be reduced for the same period of use by the appropriate selection of electrode geometry (e.g. smaller diameters, thin pins) and spark-plug concept (surface-gap spark plugs). In a surface-air-gap spark plug with four ground electrodes, there are eight possible spark gaps. In this way, the wear is evenly distributed across all four electrodes.

The electrical resistance effected in the conductive glass seal also reduces erosion and wear.

Changes in operation
Dirt and changes in the engine caused by aging (e.g. higher oil consumption) can also affect operation of the spark plug. Deposits on the spark plug can result in shunts, and thus in misfiring. This, in turn, may cause a considerable rise in pollutant emissions, and even damage the catalytic converter.

For these reasons, spark plugs have a defined service life after which the spark plugs must be replaced.

Irregular operating states
Irregular operating states may arise and the engine and the spark plugs may sustain damage due to incorrectly set ignition systems, the use of spark plugs whose heat range is unsuitable for the engine, or the use of unsuitable fuels.

Auto-ignition
Auto-ignition (pre-ignition) is an unwanted ignition process. This arises in response to overheated components (e.g. on the tip of the spark-plugs insulator nose, on an exhaust valve, on protruding cylinder-head gaskets) with the result that the air/fuel mixture ignites there under uncontrolled conditions. Severe damage can be caused to the engine and spark plug as a result of auto-ignition.

Combustion knock
Knocking is characteristic of an uncontrolled combustion process with very sharp rises in pressure (see Knocking). The combustion process is considerably faster than normal combustion. The components (cylinder head, valves, piston, and spark plugs) are subject to high temperature load which can result in damage. Knocking can be prevented by retarded ignition angle (see Knock control).

Versions and applications
The different spark-plug types are identified by a type designation. This contains all the essential spark-plug features (Figure 4, [1]). The electrode gap is additionally indicated on the packaging. The spark plug which is suitable for a given engine is specified or recommended by the engine manufacturer and by Bosch.

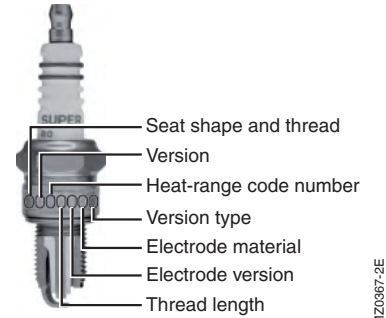
Standard spark plug for automobiles
The spark plug pictured in Figure 1 can be referred to as the standard plug for older, simply built, naturally aspirated engines.

The demands imposed on the spark plug are constantly increasing as engines are further developed up to higher specific power outputs due to, for example, improved fuel efficiency and emission-control legislation. The most sophisticated engines are direct-injection engines with multi-stage supercharging and valve-timing adjustment.

Spark plugs for direct-injection engines
Direct-injection engines impose particularly high demands on spark plugs. They must therefore be specially adapted to the needs of the respective engine (e.g. power output, mean pressure). This results in different requirements of the spark-plug concept for combustion processes with stratified or homogeneous operation.

Through the combination of direct injection with turbocharging, the emphasis in spark-plug development is on expanding

Figure 4: Designation code
The type designation describes the following features.



ignition-voltage demand, thermal load capability of the electrodes, mechanical load capability, and service life (wear reduction). For this reason, spark plugs with precious metal electrodes and M12 thread are normally used, and they are also specifically aligned in the combustion chamber when used for spray-guided combustion processes with stratified operation.

Spark plugs for gas and flex-fuel applications
The spark-plug design for natural-gas and flex-fuel use is comparable with that described previously. For each application the appropriate spark-plug version is determined while taking into account the operating conditions within the framework of the application measurements.

As a rule, alternative fuels have higher antiknock properties such that the moment of ignition can be advanced for an optimum combustion. For the most part a lower heat range is recommended for the spark plug in view of the resulting higher thermal load on the spark plug.

Special spark plugs
Special spark plugs are used for particular requirements. These differ in their design, which is dictated by the usage conditions and the installation conditions in the engine.

Spark plugs for motor sport
Engines for sports vehicles are exposed to extreme thermal loads on account of the high proportion of wide-open throttle applications involved. Spark plugs for these operating conditions mostly have precious-metal electrodes (silver, platinum) and a short insulator nose with low heat absorption.

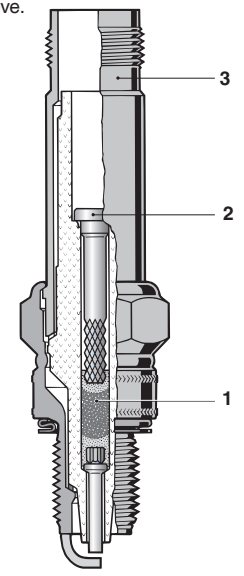
Fully-shielded spark plugs
In the event of very high interference-suppression demands, it may be necessary to shield the spark plugs. On fully-shielded spark plugs the insulator is wrapped in a metal shielding sleeve. The terminal is located inside the insulator (Figure 5). Fully-shielded spark plugs are watertight.

Spark-plug practice
Spark-plug installation
When installed correctly and the right type is chosen, the spark plug is a reliable component of the ignition system. The electrode gap can only be re-adjusted in spark plugs with front electrodes. On surface-gap and surface-air-gap spark plugs the ground electrodes must not be re-adjusted as this will alter the spark-plug concept.

Errors and their consequences
For a particular engine type only the spark plugs approved by the engine manufacturer or recommended by Bosch may be used. Using unsuitable spark-plug types can cause serious engine damage.

Incorrect heat-range code number
The heat-range code number must without fail match the engine manufacturer's spark-plug specification or the recommendation by Bosch. Auto-ignitions may result if spark plugs with a heat-range code number other than that prescribed for the engine are used.

Figure 5: Fully-shielded spark plug
1 Special conductive glass seal (interference-suppression resistor),
2 Ignition-cable terminal,
3 Shielding sleeve.



Incorrect thread length
The thread length of the spark plug must match the thread length in the cylinder head. If the thread is too long, the spark plug will project too far into the combustion chamber. Damage to the piston is a potential consequence. In addition, coked spark-plug threads may make it impossible to unscrew the spark plug or cause the spark plug to overheat.

If the thread is too short, the spark plug will not project far enough into the combustion chamber. This may result in poorer mixture ignition. Furthermore, the spark plug will not reach its self-cleaning temperature and the lower threads in the cylinder head will sustain coking.

Tampering with the seal seat
On spark plugs with conical seal seats neither a plain washer nor a sealing ring may be used. On spark plugs with flat seal seats only the "captive" sealing ring located on the spark plug may be used. It must not be removed or replaced by a plain washer. Without the sealing ring the spark plug projects too far into the combustion chamber, the transfer of heat from the spark-plug shell to the cylinder head is compromised, and the spark-plug seat provides a poor seal. If an additional sealing ring is used, the spark plug does not project far enough into the threaded hole, and the transfer of heat from the spark-plug shell to the cylinder head is likewise compromised.

Assessment of spark-plug faces
The spark-plug face refers to the combustion end of the spark plug with the electrodes and the insulator nose. Their appearance in terms of color and deposits provides indications of the spark plug's operating performance and the engine's mixture composition and combustion process.

References
[1] www.bosch-zuendkerze.de

Catalytic exhaust-gas treatment

Catalytic converter

Emission-control legislation defines limits governing pollutant emissions from motor vehicles. Engine-design measures alone are not sufficient to comply with these limits. In addition to reducing untreated engine emissions, the emphasis in spark-ignition engines is on catalytic aftertreatment of the exhaust gas in order to convert the pollutants. Catalytic converters convert the pollutants produced during combustion into harmless components.

Three-way catalytic converter

Function

The state-of-the-art technology for engines operating with a stoichiometric A/F mixture is the three-way catalytic converter (TWC). Its task is to convert pollutant components HC (hydrocarbons), CO (carbon monoxide), and NO_x (nitrogen oxides), which arise during the combustion process, into harmless components. The end products are H_2O (water vapor), CO_2 (carbon dioxide), and N_2 (nitrogen).

Design and operating principle

The catalytic converter comprises a steel-plate container as housing, the substrate and the substrate coating (washcoat) made of aluminum oxide (Al_2O_3) on which the precious metal is finely spread. The substrate is usually a ceramic monolith, though metallic monoliths are used for special applications (see Ceramic monolith and Metal monolith). On the monolith is a substrate coating, which enlarges the effective area of the catalytic converter by a factor of up to around 10,000. The catalytic layer on top of this contains the precious metals palladium or in the past platinum and rhodium. Platinum and palladium accelerate the oxidation of HC and CO, while rhodium is responsible for reducing NO_x. The amount of precious metal contained in a catalytic converter is approx. 1 to 10 g depending on the engine displacement and the exhaust emissions standard to be met.

For example, oxidation of CO and HC takes place according to the following equations:

$$2\ CO + O_2 \rightarrow 2\ CO_2,$$
$$2\ C_2H_6 + 7\ O_2 \rightarrow 4\ CO_2 + 6\ H_2O.$$

Reduction of nitrogen oxides takes place, for example, according to the following equations:

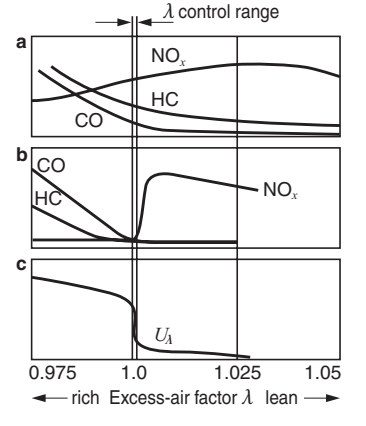
$$2\ NO + 2\ CO \rightarrow N_2 + 2\ CO_2,$$
$$2\ NO_2 + 2\ CO \rightarrow N_2 + 2\ CO_2 + O_2.$$

The oxygen required for the oxidation process is either present in the exhaust gas as exhaust-gas oxygen (resulting from incomplete combustion), or it is taken from the NO_x, which is reduced at the same time.

The concentrations of pollutants in the untreated exhaust gas (upstream of the catalytic converter) depend on the excess-air factor λ (Figure 1a). In order for the conversion of the three-way catalytic

Figure 1: Catalytic-converter efficiency as a function of the excess-air factor λ

a) Exhaust-gas emission upstream of three-way catalyst,
b) Exhaust-gas emission downstream of three-way catalyst,
c) Electrical signal of two-step λ sensor.
U_λ Sensor voltage.



converter to be as high as possible for all three pollutant components, a mixture composition in the stoichiometric ratio $\lambda = 1$ is required (Figure 1b). At $\lambda = 1$, there is a state of equilibrium between oxidation and reduction reactions which facilitates a full oxidation of HC and CO and, at the same time, reduces the NO_x. In this way, HC and CO act as reducing agents for NO_x. The "window" (λ control range) in which the mean time value of λ must lie is very small. Mixture formation must therefore be corrected in a λ control loop with the aid of the signal of a λ oxygen sensor (Figure 1c) (see λ control).

Oxygen storage
The λ accuracy in the dynamic range is typically 5 %, i.e. fluctuations of $\lambda = 1$ in this scale are inevitable. The catalytic converter can compensate for minor mixture fluctuations itself. It has the ability to store excess oxygen in the lean phase in the catalytic converter in order to release it again in the subsequent rich phase. Its substrate coating contains ceroxide, which can make oxygen available via the following balance reaction:

$$2\ Ce_2O_3 + O_2 \leftrightarrow 4\ CeO_2.$$

The task of the engine-management system is, therefore, clear: The mean time value of the resulting λ upstream of the catalytic converter must be very precise at one (a deviation of only a few thousandths is permitted). The mean value deviations, converted to oxygen input and output, must not overtax the available oxygen retained in the catalytic converter. Typical oxygen-storage values are in the range of 100 mg to 1 g; these values decrease as the catalytic converter ages. All conventional methods of diagnosing catalytic converters are based on the direct or indirect determination of this oxygen storage capacity (osc).

NO_x storage catalytic converter
Function, design and operating principle
During lean-burn operation, it is impossible for the three-way catalytic converter to convert the nitrogen oxides (NO_x) which have been generated during combustion. CO and HCs are oxidized by the high residual-oxygen content in the exhaust gas and are therefore no longer available as reducing agents for the nitrogen oxides.

The catalytic layer of the NO_x storage catalytic converter also contains substances which can store NO_x (e.g. barium oxide). All conventional NO_x storage coatings also have the properties of a three-way catalytic converter, with the result that the NO_x storage catalytic converter operates like a three-way catalytic converter at $\lambda = 1$.

In stratified lean-operation mode (see Gasoline direct injection), NO_x is not converted continuously, but instead in three stages.

NO_x storage
During storage, NO_x is initially oxidized to NO_2, which then reacts with the special oxides of the catalytic-converter surface and oxygen (O_2) to form nitrates (e.g. barium nitrate):

$$2\ NO + O_2 \rightarrow 2\ NO_2,$$
$$2\ BaO + 4\ NO_2 + O_2 \rightarrow 2\ Ba(NO_3)_2.$$

Regeneration
As the quantity of stored NO_x (load) increases, the ability to continue binding NO_x decreases. At a predefined laden state, the NO_x accumulator must be regenerated, i.e. the nitrogen oxides stored in it must be released again and converted. For this purpose, the engine switches briefly to rich homogeneous operation ($\lambda < 0.8$) to reduce NO to N_2, without emitting CO and HC in the process. In a second stage, the rhodium coating reduces the nitrogen oxides with CO:

$$Ba(NO_3)_2 + 3\ CO \rightarrow 3\ CO_2 + BaO + 2NO.$$
$$2\ NO + 2\ CO \rightarrow N_2 + 2\ CO_2.$$

The end of the storage and release phase is either calculated using a model-based method or measured with a λ oxygen sensor after the catalytic converter.

Desulfating
The sulfur contained in the fuel also reacts with the accumulator material in the catalytic layer. The result is that, over time, the amount of accumulator material available for NO_x accumulation diminishes. This creates sulfates (e.g. barium sulfate) which are very temperature-resistant and are not reduced during NO_x regeneration. For desulfating, special measures must be used to heat the catalytic converter to 600 to 650 °C before exposing it alternately to rich (λ = 0.95) and lean exhaust gas (λ = 1.05) for a few minutes. This process reduces the sulfates.

The various methods to heat up the NO_x storage catalytic converter located in the underfloor regions must be careful not to overheat the close-coupled catalytic converter.

Catalytic-converter operating temperature
Catalytic converters do not start significant conversion until they have reached a specific operating temperature (light-off temperature). On a three-way catalytic converter, this is approximately 300 °C. Ideal conditions for a high conversion rate are achieved at 400 to 800 °C.

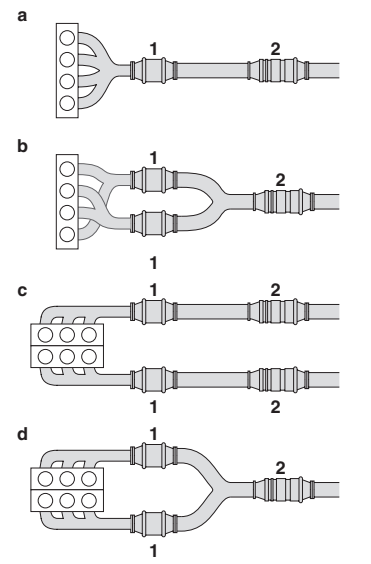
In the case of the NO_x storage catalytic converter, the favorable temperature range is lower; it reaches maximum storage capacity at 300 to 400 °C. The reason for the lower temperature range is that the maximum storage capacity is reduced at higher temperatures. At temperatures higher than 500 to 550 °C the barium compound is no longer stable, which means that nitrogen oxides can no longer be stored.

Operating temperatures of 800 °C to 1,000 °C result in intensified thermal aging of the catalytic converter. This is caused by sintering of the precious metals and the substrate coating, which reduces the active surface. At temperatures above 1,000 °C, thermal aging increases considerably until the catalytic converter has no effect at all.

Catalytic-converter configurations
The required operating temperature of the three-way catalytic converter limits the installation options. Engine-proximate catalytic converters quickly reach operating temperature, but can be exposed to very high thermal loads.

Figure 2: Catalytic-converter configurations
a) Deployment of a close-coupled catalytic converter and an underfloor catalytic converter.
b) 4-into-2 exhaust manifold for power-optimized engine design: The positioning of the underfloor catalytic converter only after the second junction is unfavorable for the heating-up characteristics, which means that preferably two close-coupled catalytic converters are deployed.
c) Engine with more than one cylinder bank (V engine): The exhaust-gas system runs completely as twin-branch with an upstream and an underfloor catalytic converter for each.
d) Engine with more than one cylinder bank (V engine): Y-shaped junction in the underfloor area to form an overall exhaust-gas train with a joint underfloor catalytic converter for both banks.

1 Close-coupled catalytic converter,
2 Underfloor catalytic converter or coated particulate filter.



A widely used configuration of the three-way catalytic converter is the split arrangement with a close-coupled catalytic converter and an underfloor catalytic converter. The close-coupled catalytic converter is optimized with regard to high-temperature stability, the underfloor catalytic converter with regard to "low light off" (low activation temperature). Figure 2 shows different possible arrangements of the primary and underfloor catalytic converters. Owing to their lower maximum permissible operating temperature, NO_x storage catalytic converters are always installed in the underfloor area.

Catalytic-converter heating

Whereas a catalytic converter at operating temperature reaches very high conversion rates of almost 100 %, considerably greater amounts of pollutants are emitted in the cold start phase and in the warm-up phase. The HC and CO emissions are particularly high when the engine is cold because fuel condenses onto the cold cylinder walls, leaves the combustion chamber unburned and cannot be converted by the cold catalytic converter.

It is therefore important to minimize untreated emissions during the warm-up phase before the catalytic converter lights off (operating temperature reached). Measures are also required to bring the catalytic converter up to operating temperature quickly. The required heat is provided by raising the exhaust-gas temperature and increasing the exhaust-gas mass flow. This is made possible by the following measures.

Ignition-angle adjustment

The main measure for increasing the exhaust-gas heat flow is ignition-angle adjustment. Combustion is initiated as late as possible and takes place during the expansion phase. The exhaust gas has a relatively high temperature at the end of the expansion phase. Retarded combustion has an unfavorable effect on engine efficiency.

Idle speed increase

A supporting measure is to raise the idle speed and thereby increase the exhaust-gas mass flow. The increased engine speed permits a greater ignition-angle retardation. However, in order to ensure reliable ignition, the ignition angles are limited to roughly 10 to 15 °KW after TDC. The heat output limited in this way is not always enough to achieve the current emission limits.

Exhaust-camshaft adjustment

A further contribution to increasing the heat flow can be achieved, if necessary, with exhaust-camshaft adjustment. The process of the exhaust valves opening as early as possible interrupts the retarded combustion early and the mechanical work generated is thus reduced further. The quantity of energy not converted into mechanical work is available as a quantity of heat in the exhaust gas.

Homogeneous split

With gasoline direct injection, there is, in principle, the possibility of multiple injection. This enables the catalytic converter to be quickly heated up to operating temperature without the need for additional components. The "homogeneous-split" measure involves initially creating a homogeneous, lean basic mixture by means of injection during the induction stroke. Subsequent injection during the compression stroke generates a stratified charge cloud. This facilitates retarded moments of ignition and results in high exhaust-gas heat flows. The exhaust-gas heat flows that can be achieved are comparable with those of a secondary-air injection.

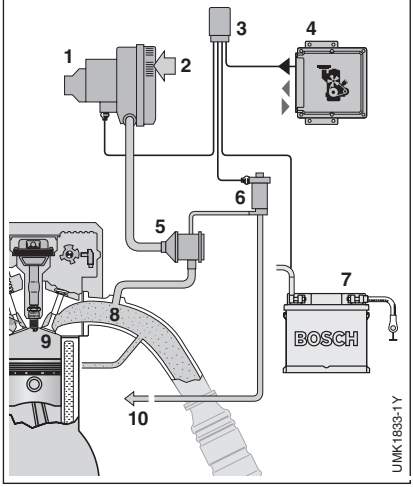
Secondary-air injection

Thermal afterburning of unburnt fuel constituents increases the temperature in the exhaust system. A rich (λ = 0.9) to very rich (λ = 0.6) air/fuel basic mixture is adjusted for this purpose. A secondary-air pump supplies oxygen to the exhaust system (Figure 3) to produce a leaner exhaust-gas composition. Where the basic mixture is very rich (λ = 0.6), the unburnt fuel constituents oxidize above a specific temperature threshold exothermically before entering the catalytic converter. To

achieve this temperature, it is necessary on the one hand to raise the temperature level with retarded ignition angles and on the other hand to introduce the secondary air as closely as possible to the exhaust valves. The exothermic reaction in the exhaust system increases the heat flow to the catalytic converter and therefore shortens its heating period. The HC and CO emissions are for the most part reduced before entering the catalytic converter.

Where the basic mixture is less rich (λ = 0.9), there is no significant reaction ahead of the catalytic converter. The unburnt fuel constituents oxidize in the catalytic converter and heat it up from the inside. For this purpose, however, the end face of the catalytic converter must first be brought up to a level above the light-off temperature by means of conventional measures (e.g. ignition-angle adjustment).

Figure 3: Secondary-air system
1 Secondary-air pump,
2 Inducted air,
3 Relay,
4 Engine control unit,
5 Secondary-air valve,
6 Control valve,
7 Battery,
8 Point of introduction into exhaust pipe,
9 Exhaust valve,
10 To intake-manifold connection.



As a rule, a less rich basic mixture is adjusted, because, if the basic mixture is very rich, the exothermic reaction ahead of the catalytic converter occurs reliably only under stable boundary conditions.

Secondary-air injection is performed with an electric secondary-air pump, which is switched by means of a relay on account of the high power demand. Since the secondary-air valve prevents backflow of exhaust gases into the pump, it must remain closed when the pump is deactivated. Either it is a passive non-return valve or it is actuated by purely electrical or (as shown in Figure 3) pneumatic means with an electrically actuated control valve. When the control valve is actuated, the secondary-air valve opens in response to the intake-manifold vacuum. The secondary-air system is coordinated by the engine ECU.

Alternative concepts for active heating
As a supplement in special cases, an electrically heated catalytic converter is used for quick heating up of the catalytic converter. This has been used previously in individual small-series projects.

λ control

To insure that the conversion rates of the three-way catalytic converter are as high as possible for the pollutant components HC, CO and NO_x, the reaction components must be present in the stoichiometric ratio. This requires a mixture composition of $\lambda = 1.0$, i.e. the stoichiometric air/fuel ratio must be adhered to very precisely.

Mixture formation must be followed up in a control loop, because sufficient accuracy cannot be achieved solely by controlling the metering of the fuel. When the λ control loop is used, deviations from a specific air/fuel ratio can be detected and corrected through the quantity of fuel injected. The residual-oxygen content in the exhaust gas, which is measured with λ oxygen sensors (see Two-step λ oxygen sensor and Broad-band λ oxygen sensor), serves as the measure for the composition of the air/fuel mixture.

Two-step λ control

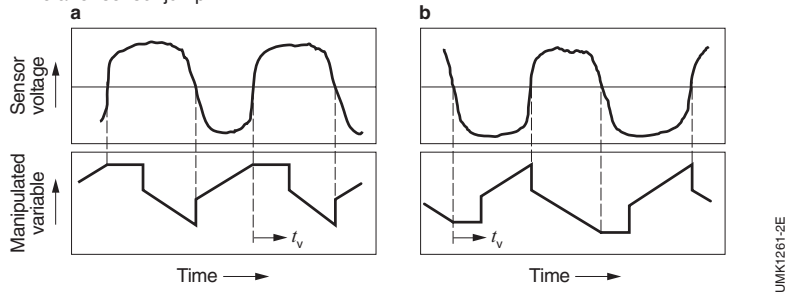
Two-step λ control adjusts the mixture to $\lambda = 1$. A manipulated variable, composed of the voltage jump and the ramp, changes its direction of control for each voltage jump of the two-step λ sensor. This indicates a change from rich to lean or from lean to rich (Figure 4). The typical amplitude of this manipulated variable has been set in the range of 2 to 3 %. The result is a limited controller dynamic, which is predominantly determined by the sum of the response times (pre-storage of fuel in the intake manifold, four-stroke principle of the spark-ignition (SI) engine, and gas travel time).

The typical shift of the oxygen zero transition point (theoretically at $\lambda = 1.0$) and therefore the jump of the λ sensor, caused by the variation in exhaust-gas composition, can be control-compensated by shaping the manipulated variable's characteristic curve asymmetrically (rich or lean shift). Here, the preferred method is to hold the ramp value at the voltage jump for a controlled dwell time t_V after the sensor is subjected to a voltage jump. A rich or lean shift is performed depending on the operating point. When a shift is made towards "rich", the manipulated variable holds at the rich position for a dwell time t_V although the sensor signal has already jumped towards "rich" (Figure 4a). Jump and ramp of the manipulated variable do not also move towards "lean" until the dwell time has elapsed. If the sensor signal then jumps towards "lean", the manipulated variable directly counteracts this (with jump and ramp) without dwelling on the lean position. The dwell time t_V is the result of a pre-control coefficient, which is corrected by a proportion from the two-sensor λ control system (post cat control).

The behavior is reversed in a shift to "lean". If the sensor signal indicates a lean mixture, the manipulated variable holds at the lean position for the dwell time t_V (Figure 4b) and then adjusts towards "rich". A

Figure 4: Manipulated-variable curve of a two-step control with a λ sensor upstream of the catalytic converter and λ shift (delay time t_v) due to pre-control proportion and post cat control

a) Shift to "rich", b) Shift to "lean".
t_v Dwell time after sensor jump.



sensor signal jump from "lean" to "rich" is, however, immediately counteracted.

Continuous λ control

The defined dynamic response of a two-step control can only be improved if the deviation from $\lambda = 1$ can actually be measured. The broad-band λ sensor can be used to achieve continuous-action control at $\lambda = 1$ with a stationary, very low amplitude in conjunction with high dynamic response. The control parameters are calculated and adapted as a function of the engine's operating points. Above all, with this type of λ control, compensation for the unavoidable offset of the stationary and non-stationary pilot control is far quicker.

The broad-band λ oxygen sensor also enables adjustment to mixture compositions that deviate from $\lambda = 1$. Thus, controlled enrichment ($\lambda < 1$), e.g. for component protection, can be effected. The broad-band λ sensor also enables controlled leaning ($\lambda > 1$), e.g. for a leaner warm-up during catalytic-converter heating.

Two-sensor λ control

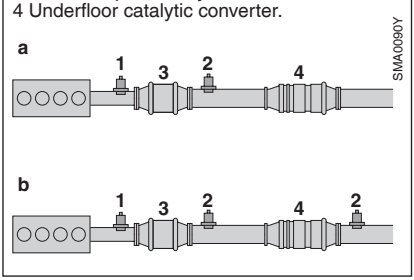
When it is situated upstream of the catalytic converter, the λ sensor is heavily stressed by high temperatures and untreated exhaust gas, and this leads to limitations in λ sensor accuracy. The voltage-jump point of a two-step sensor or the characteristic curve of a broad-band sensor can shift, for example, as a result of changed exhaust-gas compositions. A λ sensor downstream of the catalytic converter is subjected to these influences to a much lesser extent. However, λ control on its own with the sensor downstream of the catalytic converter demonstrates disadvantages in dynamic response on account of the gas travel times, and responds to mixture changes more slowly.

Greater accuracy is achieved with two-sensor control. Here, a slower correction control loop is superimposed on the two-step or continuous-action λ control described by means of an additional two-step λ sensor (Figure 5a). The voltage of the two-step sensor downstream of the catalytic converter is compared with a setpoint value (e.g. 600 mV) for this purpose. On this basis the control evaluates the deviations from the setpoint value and additionally alters over a dwell time t_v additively the controlled rich or lean shift of the first control loop of a two-step control or the setpoint value of a continuous-action control.

Three-sensor λ control

There are SULEV concepts (Super Ultra-Low-Emission Vehicle) with a third sensor downstream of the underfloor catalytic converter. The two-sensor control system (single cascade) has been supplemented here by an extremely slow closed-loop control system using a third sensor located downstream of the underfloor catalytic converter (Figure 5b). The aim of development is, however, the two-sensor concept.

Figure 5: Installation locations of λ sensors
a) Two-sensor control,
b) Three-sensor control.
1 Two-step or broad-band λ sensor,
2 Two-step λ sensor,
3 Close-coupled catalytic converter,
4 Underfloor catalytic converter.



Particulate filter for gasoline engines

Function

The function of the particulate filter in the gasoline engine with gasoline direct injection is to adhere to the limit values for particulate matter and particulate number laid down in the exhaust-emissions standard Euro 6c or Euro 6d(-temp) (see also Particulate filter for diesel engines). An additional catalytic coating of the particulate filter serves to increase the catalytic-converter volume available for the oxidation of CO and HC and the reduction of NO_x. A catalytically coated particulate filter can thus replace a three-way catalytic converter.

Design

The particulate filter for gasoline engines is designed in the same way as the particulate filter in diesel engines (see Particulate filter for diesel engines). For gasoline applications the ceramic is currently made of cordierite.

Regeneration of the particulate filter

Soot is oxidized by oxygen only at temperatures above approximately 550 to 600 °C. In gasoline applications of the particulate filter the temperatures reached in normal operation and the requirement of additional engine measures to increase the exhaust-gas temperature are determined by the filter installation position. Usually the filter is positioned downstream of the three-way catalytic converter. When installed directly downstream of the three-way catalytic converter temperatures of 550 to 600 °C can be achieved in part-load operation without additional measures, while this is only achieved with an installation position in the underfloor area in high-load operation. In terms of regeneration performance an installation position directly downstream of the three-way catalytic converter therefore exhibits clear advantages, but the space required for this must be available near the engine. Integrating the filter function into the three-way catalytic converter for the four-way catalytic converter (FWC) does exhibit clear advantages with regard to the regeneration performance and near-engine installation area, but up to now has only been implemented in individual applications due to the highest thermal load at full load.

The most important engine measure to increase the exhaust-gas temperature is, as with heating the catalytic converter, ignition-angle retardation, which simultaneously brings about de-throttling and an increase in the exhaust-gas mass flow and in the case of gasoline direct injection can be combined with multiple injection ("homogeneous split"). Another measures is the alternate rich/lean adjustment of individual cylinders while adhering to an exhaust-gas λ of 1. Unburned exhaust-gas constituents from the rich-burn cylinders are exothermally converted here with O_2 from the lean-burn cylinders in the three-way catalytic converter.

In contrast a diesel engine, in a gasoline engine operated with exhaust-gas λ of 1 there is no O_2 available for soot oxidation in normal operation. After the required temperature is reached, the soot can therefore be burned off initially in an overrun phase. Alternatively, it is possible to achieve soot burn-off by setting a lean exhaust-gas λ of up to approximately 1.1, which is however to be avoided because of the increased NO_x emissions involved.

Saturation detection

The strategies used to determine the amount of soot stored in the particulate filter are comparable with the method used on diesel engines (see Particulate filter for diesel engines).

Motor sport

Requirements

The basic requirements regarding components in motor-sport vehicles are different from those in production vehicles. A racing car requires maximum performance for a few hours with the lowest possible weight, whereas durability and low costs are the desired attributes in a production car. Furthermore, the components in a racing car are exposed to greater loads from heat, dust, moisture, and vibrations than those in a production car.

The Bosch Motorsport Division develops, produces and sells electronic systems and components for this special field of application. Both modified mass-produced parts and pure motor-sport developments, which often have no equivalent in series production, are offered.

Components

Modified mass-produced parts

Reducing weight

Technicians are constantly on the lookout for ways of reducing the weight of production components for racing applications. The reduced service life generally associated with this is tolerated.

By using ultralight special materials and specific special constructions, it is possible to reduce the weight of many components – and all this with higher power output. As the motto goes: Remove as much material as possible, keep as little material as is necessary. A quotation that goes around in design-engineer circles puts it in a nutshell: "The perfect racing car crosses the finishing line in first place and then falls apart into its component parts".

Increasing performance

There is a huge range of measures to increase the performance of production components. Alternators for instance have additional windings and thereby produce more energy. Their rotors are fine-balanced and are those less sensi-

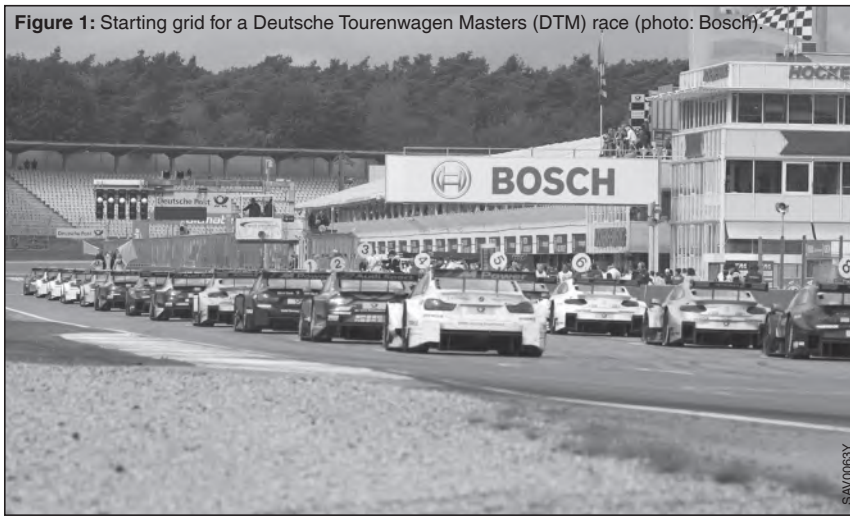


Figure 1: Starting grid for a Deutsche Tourenwagen Masters (DTM) race (photo: Bosch).

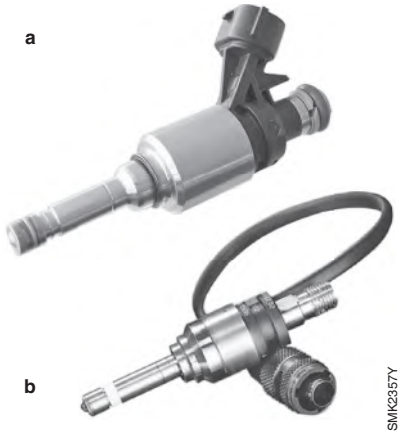
tive to high rotational speeds. Temperature sensors respond much more quickly once the metal that normally encloses the measuring capsule has been removed. Changes to fuel pumps and fuel injectors are frequently made with increased throughflow and increased pressure in mind.

The sometimes significantly higher specific engine performance of racing cars makes these measures absolutely necessary. The use of better ferromagnetic materials generate higher magnetic forces in the coils of the fuel injectors. This makes possible greater needle strokes, which in turn permit a higher fuel flow.

The modifications to ignition coils go even further. Here only the original shell remains the same, the technology inside is all new. The coils are made with fewer windings from thicker wire, the coil cores are made of higher-quality material. The higher spark energy required for direction-injection highly supercharged racing engines can be guaranteed in this way. The shorter dwell periods increase the upper speed limit to 15,000 rpm.

Figure 2: High-pressure fuel injectors in motor sport

a) Component with production connector on the housing.
b) Component with cable and motor-sport connector.



Replacing connectors

Production connectors made of plastic are not designed for frequent opening and closing and have only limited resistance to vibrations and moisture. They are therefore often removed from motor-sport components and replaced by cables with motor-sport connectors (Figure 2). Their metal housings have a bayonet catch for reliable and rapid opening and closing. They are dust-proof and watertight and, because they are no longer rigidly connected to the component, can be installed in locations that are less exposed to vibration. Particularly for sensors, ignition coils and fuel injectors this modification makes installation and removal easier and increases operational reliability by minimizing primarily vibration-induced damage. These connectors are standard in motor racing and are compatible with virtually all wiring harnesses and diagnostic tools.

Dissipating heat

High-performance electronic systems generate a lot of heat which can jeopardize temperature-sensitive components. This heat is dissipated by cooling fins to the housing walls and the use of thermal-conduction paste in the housings. These housings are often black to improve heat dissipation.

Regular maintenance

Unlike in production-vehicle manufacturing, where starters and alternators as a rule go through the entire life of the vehicle without maintenance and repairs, these components are overhauled in motor racing after a certain period of operation. Bosch offers its customers this service for all motor-sport components. The components are dismantled, cleaned and provided with new wearing parts at the factory. They are then restored to the same operational reliability as new parts.

ECUs

The fundamental difference from a production ECU, where the manufacturer must guarantee tuning protection measures, is the free access to the software for all motor-sport systems. In this way, the teams can tune each car individually to racetrack, weather and driver. The computer programs required for tuning on the laptop computer are developed by Bosch itself. Furthermore, programs for analyzing recorded vehicle data and for simulating different vehicle tunings are offered.

Series-based ECUs
Series-based developments with sheet-metal housings and plastic connectors constitute the entry level into motor-sport ECUs. The range of functions is less complex than in high-end ECUs and can therefore also be easily handled by smaller teams. These systems are frequently used in Formula 3, in marque cups and in near-series motor-racing classes, where many GT3 class customer sports cars are used.

High-end ECUs
Top-end models among the ECUs have all-aluminum milled housings and motor-sport connectors (Figure 3). The housings are powder-coated to protect the surface. The number of connector pins and hence of inputs and outputs is high, and the number of controllable cylinders is frequently high too. Some systems have customer code areas, in which customers can store their own software and link with the basic software. Communication channels to CAN and LIN bus systems as well as telemetry connections and internal data memories are standard.

Working with these systems calls for a high level of specialized knowledge, which as a rule can only be brought to bear by works teams. Typical areas of application for these ECUs are the World Rally Championship (WRC), the World Endurance Championship (WEC), and the Deutsche Tourenwagen Masters (DTM).

Pure motor-sport developments

As well as the high-end ECUs, further pure motor-sport developments are featured in the Bosch Motorsport portfolio. These include, for example, lap-trigger systems with which the times of complete laps or lap segments are measured and telemetry systems for data transfer from the racing car to the box. Driver displays replace all the instruments in the dashboard, are freely programmable, and have switchable display pages. They are frequently equipped with data memories.

High-pressure fuel-injection systems and their valve injectors often require a significantly higher operating voltage than the conventional vehicle electrical system can provide. This gap is filled by the power supply unit which is used in many Formula 1 cars and delivers voltages of up to 90 V.

Safety and comfort
Electronic systems from Bosch however not only deliver better acceleration, but also better deceleration and increased safety and comfort in racing cars.

The Bosch motor-sport ABS systems come with different control characteristics which the driver can select by means of a rotary switch in the dashboard. Here too the ECU software is open and can be changed where necessary. Pre-parameterized plug-and-play versions are offered for certain cup cars. The experts from Bosch lend the customers their sup-

Figure 3: High-end ECU with milled housing and motor-sport connectors



port during individual tuning of the ABS systems on the test bench or on the racetrack.

A radar-based accident-avoidance system uses visual signals to warn against passing cars; these systems are particularly helpful when it is raining or at night.

The Bosch Power-Box replaces all the relays and fuses in a racing car and enables the relay switching points to be changed on a laptop computer without a mechanic having to undo a screw or exchange a component. The laptop computer here also replaces the tool box.

In the case of systematic troubleshooting it is very helpful to be able to measure specifically the signal flow of individual cables in the wiring harness. Here a specifically developed breakout panel, which is switched between the wiring harness and the ECU, provides sterling service.

Figure 4: Spark-plug variants for racing use
a) Racing spark plug with 8 mm thread, toothing for special tool and platinum electrode.
b) Stripped-back spark plug for subsequent processing by the motor-racing customer.



Customer-specific extra-small series

Bosch hand-makes in small numbers motor-sport wiring harnesses which network the systems in a racing car with each other. They have the customary motor-sport connectors and are delivered to the customer after exhaustive functional tests.

Motor-sport spark plugs only have the basic idea in common with the mass-produced components. Electrodes made of pure platinum permit operation at significantly higher temperatures. Thread diameters reduced to 8 mm allow the engine developers plenty of freedom in fashioning the combustion chambers and the cooling ducts (Figure 4). Thread length, width across flats and other sizes are designed by arrangement with the customer. This results in hand-made sample spark plugs which are specially customized to the existing engine concept.

As well as the ready-to-install spark plugs, Bosch offers its motor-racing customers the separate ceramic core of the spark plug for subsequent processing. The customer provides this “stripped-back spark plug” with their own shell, the design of which is among the best-kept secrets of the design engineers.

New directions

Mirroring developments in production-vehicle manufacturing, hybrid systems are increasingly being used in racing cars. Electric motors work in coordination with gasoline or diesel engines; recovered energy is buffered in electrical or mechanical accumulators.

Some motor-racing classes are already fully reliant on the purely electric drivetrain, as is demonstrated for example by Formula E cars, which race on city-center street circuits around the world.

In these new directions too the long tradition of exchanging experience and technology between series production and motor sport is being continued by Bosch.

LPG operation

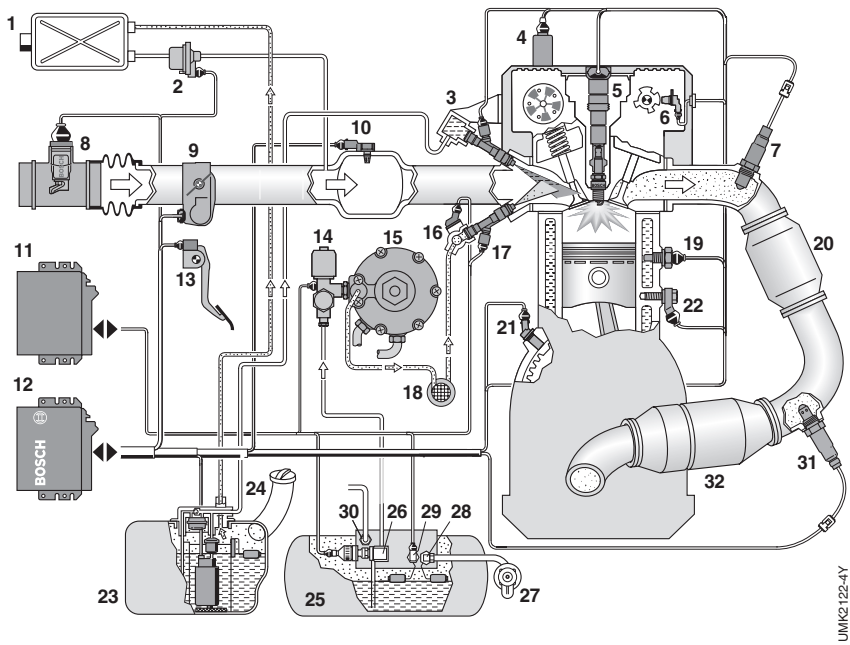
Applications

Liquified petroleum gas (LPG) is obtained during the extraction of crude oil and natural gas and during different refinery processes. It is made up primarily of propane and butane (see Liquid petroleum gas).

As the conventional fuels gasoline and diesel have become more expensive in the last few years, the cheaper liquified petroleum gas – not least due to regional tax reductions (e.g. for Germany, Italy, Poland) – is becoming increasingly more widespread. Liquified petroleum gas is widely used in Europe and in South Korea, Australia, and China. At present (as at No-

Figure 1: Bivalent gasoline-engine operation with gaseous LPG injection

1 Activated charcoal canister, 2 Canister-purge valve, 3 Gasoline rail with gasoline injectors,
4 Camshaft adjuster, 5 Ignition coil with spark plug, 6 Camshaft sensor,
7 λ sensor upstream of primary catalytic converter, 8 Hot-film air-mass meter,
9 Electronically actuated throttle valve (ETC), 10 Intake-manifold pressure sensor,
11 LPG ECU, 12 Engine ECU, 13 Accelerator-pedal module,
14 LPG shutoff valve, 15 Evaporator with pressure regulator,
16 Combined pressure and temperature sensor, 17 LPG rail with gas injectors,
18 LPG filter, 19 Engine-temperature sensor, 20 Primary catalytic converter,
21 Crankshaft speed sensor, 22 Knock sensor,
23 Gasoline tank with integrated electric fuel pump, 24 Gasoline filler neck,
25 LPG tank (steel tank), 26 Electromagnetic shutoff valve,
27 LPG filler neck, 28 80% filler stop valve, 29 LPG level indicator,
30 Pressure-relief valve, 31 λ sensor downstream of primary catalytic converter,
32 Main catalytic converter.



vember 2016), there is estimated to be a population of more than 26 million LPG-fueled vehicles worldwide [1]. Up-to-date information on the number of LPG-fueled vehicles and the LPG filling-station network in Germany (currently around 75,350 filling stations worldwide, around 7,050 filling stations in Germany) can be found on the Internet (see [2] and [3]).

The additional volumetric consumption of LPG-fueled vehicles is, depending on engine concept, engine management and gas system 20 – 30 %, compared with gasoline-fueled vehicles. In spite of this additional costs, the running costs of LPG vehicles are cheaper than those of gasoline and diesel vehicles.

LPG systems are distinguished by their mixture formation with gaseous injection or liquid LPG injection and by the ECU concepts with one common or two separate engine ECUs (master/slave concept). Systems with venturi mixers (operating according to the carburetor principle) can only still be used in older vehicles or in markets with laxer exhaust-emission regulations.

Vehicles with spark-ignition engines are virtually exclusively converted to bifuel LPG/gasoline operation. Most LPG systems are installed as a retrofit. In the meantime, however, vehicle manufacturers are also offering models with bifuel LPG/gasoline operation off the assembly line. In these vehicles, the LPG tank – as in the retrofit concepts – is installed for the most concepts in the spare-wheel compartment. The luggage-compartment load volume remains unchanged with this form of installation. The spare wheel is simply omitted and replaced by a blow-out repair kit (Tire-Fit).

Design

Storage of LPG

LPG can liquify at ambient temperatures of – 20 to +40 °C under pressure (3 – 15 bar), depending on the propane-to-butane mixture ratio. It is stored in a liquid state in steel tanks. The filling system of the LPG tanks incorporates a fill-volume restrictor which ensures that the fill volume does not exceed 80 % of the liquid phase of the maximum tank volume. This ensures that the internal tank pressure does not rise to unacceptable levels when the tank is subjected to heat. The tanks are subject (also in some non-European countries) to European safety regulations governing their properties, equipment specification and installation – currently ECE-R67-01 [4].

The safety devices of an LPG tank comprise a pressure-relief valve, an electromagnetic shutoff valve with integrated flow limiter (pipe-fracture protection), and optionally a thermal safety valve which allows the gas to escape under controlled conditions in the event of a fire.

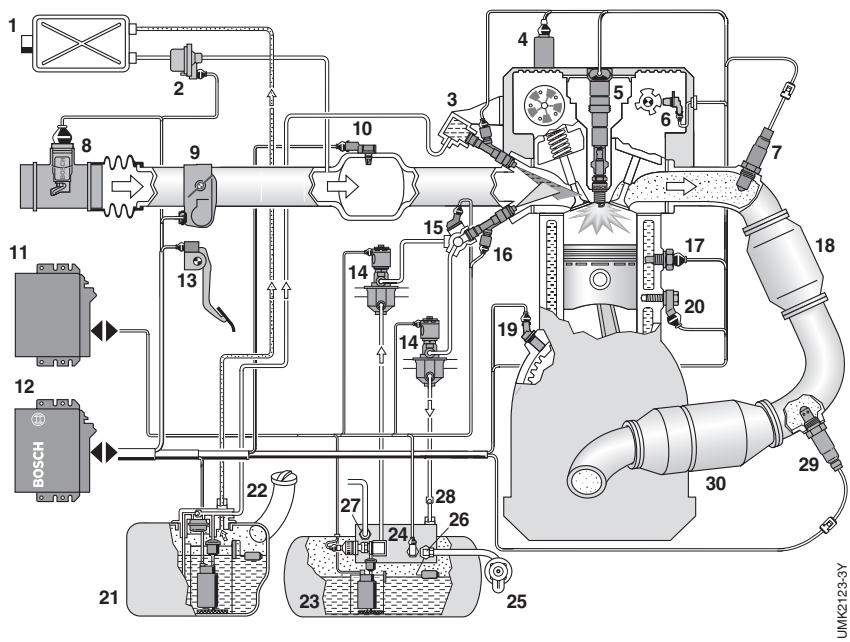
Fuel delivery

The LPG is delivered from the tank to the engine in systems with gaseous injection by the internal tank pressure (LPG vapor pressure). Because this pressure drops as ambient temperature decreases, the operation of such systems is partly limited with regard to their suitability for winter operation.

In systems with liquid injection, the LPG is delivered by an LPG fuel-supply pump installed in the LPG tank which generates a delivery pressure in addition to the internal tank pressure. In terms of their suitability for winter operation, these systems are comparable with gasoline injection systems.

Components
The following components are integrated into LPG vehicles (Figures 1 and 2):
– LPG filler neck with integrated non-return valve,
– Steel tank (cylindrical or toroidal),
– Pressure-relief valve (27 bar),
– LPG level sensor,
– Valve with integrated 80 % tank filler stop,
– Electromagnetic shutoff valve (in the tank),
– LPG shutoff valve (in the engine compartment),
– Evaporator with pressure regulator (only for systems with gaseous LPG injection),
– Pressure regulator (optional in systems with liquid LPG injection),
– LPG fuel rail with gas injectors,
– Pressure and temperature sensors (on the rail),
– LPG filter,
– If necessary, a slave ECU,
– Demand-controlled LPG supply pump (only in systems with liquid LPG injection).

Figure 2: Bivalent gasoline-engine operation with liquid LPG fuel injection into the intake manifold
1 Activated charcoal canister, 2 Canister-purge valve, 3 Gasoline rail with gasoline injectors,
4 Camshaft adjuster, 5 Ignition coil with spark plug, 6 Camshaft sensor,
7 λ sensor upstream of primary catalytic converter, 8 Hot-film air-mass meter,
9 Electronically actuated throttle valve (ETC), 10 Intake-manifold pressure sensor,
11 LPG ECU, 12 Engine ECU, 13 Accelerator-pedal module,
14 LPG shutoff valves for supply and return, 15 Combined pressure and temperature sensor,
16 LPG fuel rail with gas injectors, 17 Engine-temperature sensor,
18 Primary catalytic converter, 19 Crankshaft-speed sensor, 20 Knock sensor,
21 Gasoline tank with integrated electric fuel pump, 22 Gasoline filler neck,
23 LPG tank with integrated electric LPG pump, 24 Electromagnetic shutoff valve,
25 LPG filler neck, 26 80 % filler stop valve with integrated LPG level sensor,
27 Pressure-relief valve, 28 Non-return valve,
29 λ sensor downstream of primary catalytic converter, 30 Main catalytic converter.



LPG systems

Systems with gaseous LPG injection

Operating principle

The LPG is delivered in liquid state from the tank via the electromagnetic shutoff valve to the engine compartment (Figure 1), where an additional shutoff valve is installed. After this, the LPG is converted into the gaseous state with pressures of 0.5 to 3.5 bar in the evaporator and the pressure regulator. The evaporator is heated by the engine coolant in order to compensate for the cooling encountered during the phase transition. The gaseous LPG is directed from the evaporator through a pressure-resistant flexible hose to the LPG rail and from here injected by the gas injectors into the intake manifold.

The air-intake section of the basic engine remains unchanged except for the connection of the LPG gas injectors to the intake manifold.

Mixture formation

Mixture formation is effected in the same manner as conventional multipoint gasoline injection. The LPG rail supplies the individual gas injectors with gaseous LPG. Engine management determines the required injected fuel quantity, which is corrected by means of λ control. The injected fuel quantity is additionally corrected by means of the pressure and the temperature of the fuel in the LPG rail.

LPG injection displaces some of the inducted air volume in the intake manifold, which results in charge losses in non-turbocharged engines. Engine performance reduced by a factor of 2–4% compared with gasoline operation can thus be expected for LPG operation.

Systems with LPG liquid injection into the intake manifold

Operating principle

In systems with LPG liquid injection into the intake manifold (Figure 2) the LPG ECU regulates the system pressure via a fuel pump installed in the LPG tank.

Some systems may also need a pressure regulator to regulate the injection pressure. Because of the tendency for vapor bubbles to form in the fuel system, systems with LPG liquid injection must be fitted with a tank return line in order if necessary to expel vapor bubbles that occur.

Mixture formation

As with conventional multipoint gasoline injection, the LPG is injected into the intake manifold in front of the engine intake valves. The evaporation that occurs as the LPG is injected (according to the Joule-Thomson effect) lowers the temperature of the inducted air/gas mixture, which has a positive impact on the engine's volumetric efficiency and performance figures.

Advantages of LPG liquid injection

LPG liquid injection offers the following advantages:

- Gas operation is also possible at extremely low temperatures.
- Monofuel gas operation is possible.
- Engine start in LPG mode is also possible in bifuel systems.
- A better cylinder charge is obtained thanks to the cooling effect in the intake manifold and combustion chamber, thus producing identical engine-performance figures to gasoline without further measures. Engines with LPG liquid injection can even deliver slightly higher performance figures when compared with gasoline operation.

Systems with direct LPG liquid injection into the combustion chamber

LPG behaves very similar to gasoline at higher system pressures. Therefore LPG direct injection (LPG-DI) into the combustion chamber is also possible. LPG direct injection offers some significant advantages:

- A better exhaust-gas composition is obtained (virtually no particulate emissions) with comparable engine-performance figures.
- Very good over all thermodynamic behavior.

LPG direct liquid injection into the combustion chamber is possible in all modern gasoline engines with direct injection. Such systems are currently still in the development phase (see e.g. [5]).

Exhaust emissions

With regard to the limited pollutant components HC, CO and NO_x, the use of LPG as fuel in modern spark-ignition engines offers some advantages over gasoline operation. The situation when it comes to CO_2 emissions is even better. Here, LPG vehicles can achieve a reduction of approximately 10 %. This reduction is founded on the lower carbon content in LPG. Furthermore, when compared with diesel engines, there are virtually no particulates. LPG drives also have advantages when it comes to other exhaust-gas components which are not limited by law but which are harmful to health. Thus, for instance, LPG drives emit significantly lower amounts of aromatic hydrocarbons (e.g. benzene).

Unlike engines fueled by CNG (Compressed Natural Gas), engines fueled by LPG are not subject to any particular demands on the exhaust-gas treatment system of the existing gasoline concept. The essential exhaust-gas components of LPG operation are comparable with those of gasoline operation.

Engine management for LPG systems

Both double-ECU concepts (one ECU for gasoline operation and one for LPG operation) and single-ECU concepts (gasoline and gas functions integrated into a single ECU) are used. Because the vehicles are bifuel, the driver can select between gasoline and LPG operation using a switch.

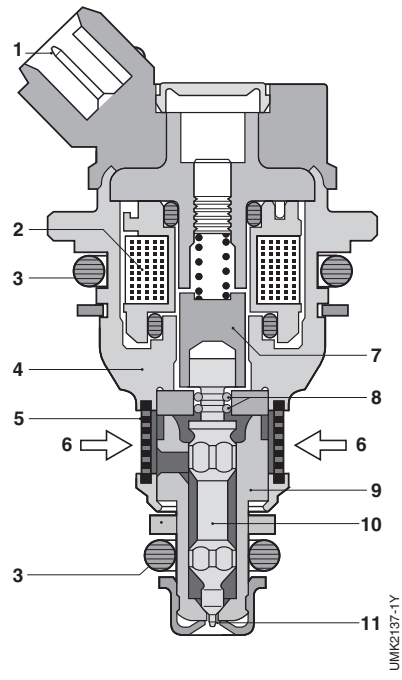
The level sensor installed in the LPG tank signals the current LPG tank level to the engine-management system. A combined low-pressure and temperature sensor mounted on the gas rail enables the engine-management system to correct the injection timing in such a way that a stoichiometric mixture can be pilot-controlled in the intake manifold even if the gas density fluctuates. The engine-management systems uses the values from the λ sensor and different adaptation algorithms to adapt the mixture formation for different gas qualities. Even functions for switching between gas and gasoline operation are integrated.

The other sensors and actuators of the engine-management system are for the most part identical to those in a gasoline engine.

Systems with LPG liquid injection into the combustion chamber require, in addition to the functions for mixture formation and gas-quality adaptation, functions for controlling the LPG fuel pump.

Figure 3: Electromagnetic liquid LPG injector

1 Electrical connection,
2 Field winding,
3 O-rings,
4 Valve housing,
5 Filter strainer,
6 LPG supply,
7 Armature,
8 O-rings,
9 Valve body,
10 Valve needle,
11 Injection nozzle.



Components

LPG gaseous injector

Function and requirements

When LPG is injected in the gaseous state of aggregation, the given energy density and the low gas pressures resulting from the state of aggregation of approx. 0.5 to 3.5 bar dictate that the following parameters must be taken into account in the design of LPG injection valves:

– Configuration of the injection valves for large gas volumes,
– Adaptation of the sealing mechanisms and materials to the gaseous medium,
– Consideration of potentially aggressive fuel contaminants in the choice of material.

Design and operating principle

Essentially, the design and function of a gas injector configured for operation with gaseous LPG are very similar to those of a CNG injector (see Engines fueled by natural gas).

LPG liquid injector

Function and requirements

In systems where the LPG is injected in the liquid state of aggregation the LPG must be kept in the liquid state in the entire pipe and rail system by means of pressure build-up by a pump installed in the LPG tank. Only in this state can it be injected.

LPG injectors must satisfy the following requirements:

– Operation at variable pressures of 2 to 25 bar.
– Absolute tightness in the case of systemic pressure variability (e.g. on shutdown of gas operation and spontaneously occurring pressure peaks due to the LPG evaporating in the rail system).
– Possibility of a flushing process to eliminate vapor bubbles that occur (e.g. during hot start or switching from gasoline to LPG operation with the engine at operating temperature).

Design and operating principle

LPG injectors (Figure 3) are, in response to the above-mentioned requirements, designed according to the bottom-feed principle. The injector is designed in such a way that the fitting for the fuel inlet into the injector housing is as close as possible to the injector outlet. This layout enables the valve to be easily flushed with fresh fuel in order to quickly remove possible vapor bubbles.

The fuel injectors normally used in gasoline engines, on the other hand, operate for the most part according to the top-feed principle. Because of the supply of fresh fuel located at the top end of the injector housing, the flush process for the purpose of removing vapor bubbles in the injector would be very diffecult or not possible at all. Under such conditions it would not be possible to provide a correct metering of the fuel quantity and thus optimum combustion, which is why the top-feed principle is not suitable for use in liquid LPG injection.

Installation position

The position of the LPG injectors in the intake manifold is subject to special design requirements according to the generation of cold by the evaporation of the LPG during the injection process (as a result of the Joule-Thompson effect). The liquid LPG evaporates abruptly on entry into the intake manifold and extracts the required energy in the form of heat out of the fresh air in the intake manifold. This specific-point cold in the intake manifold can cause unwelcome operating phenomena (e.g. misfires), which can be put down among other things to the formation of ice crystals.

LPG filler neck

Neck variants

The filler neck is not internationally standardized. For this reason, different fueling adapters are required from country to country. Four different types of LPG fuel nozzle are used in filling stations. Different filler necks are therefore needed depending on the field of application. The following necks are widely used:

– ACME neck (named for the American trapezoidal-thread standardization) is used in Germany, Belgium, Ireland, England, Luxembourg, Scotland, Austria, China, Canada, USA, Australia, and Switzerland.
– DISH neck (named for its dish shape) is used in Denmark, France, Poland, Greece, Hungary, Italy, Austria, Portugal, the Czech Republic, Turkey, China, Korea, and Switzerland.
– Bayonet neck is used in the Netherlands, Belgium and the UK.
– Euronozzle is currently used in Spain only.

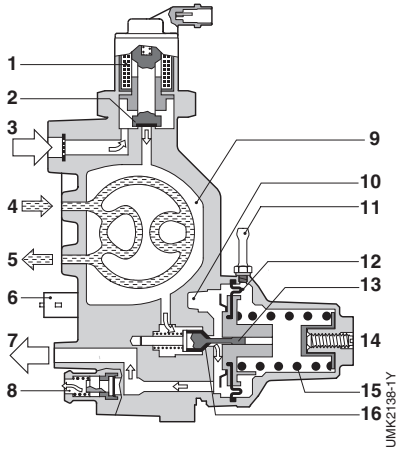
The Euronozzle is supposed to replace all the other necks on a Europewide basis in future. The EU Regulation for this standardized neck has not yet been adopted by the majority of EU member states.

Design

Each filler neck in the vehicle is designed in such a way that the LPG vehicle can be refueled at all LPG filling stations by using a suitable adapter.

Figure 4: Evaporator with pressure regulator

1 Solenoid coil of electromagnetic shutoff valve,
2 Seal,
3 Inlet for liquid LPG and filter,
4 Inlet for engine coolant,
5 Outlet for engine coolant,
6 Temperature sensor,
7 Outlet for gaseous LPG at constant pressure,
8 Safety pressure-relief valve,
9 Evaporation chamber,
10 Area for gaseous LPG,
11 Connection for intake-manifold vacuum,
12 Pressure diaphragm with mounting,
13 Control rod,
14 Adjusting screw,
15 Spring,
16 Control plunger and throttling orifice.



Evaporator for systems with gaseous LPG injection

Function

Evaporators are used in systems with gaseous LPG injection. The LPG is delivered to the evaporator by the vapor pressure in the gas tank via an electromagnetic shutoff valve.

The evaporator is one of the main components of every LPG injection system and performs two functions: Firstly, the liquid gas is converted from the liquid to the gaseous state in the first part of the evaporator. Secondly, the pressure of the gaseous LPG is then regulated to approx. 1 bar differential pressure in relation to the intake manifold. This pressure, with which the injection valves are supplied, must be kept constant through all the operating states and the varying vapor pressures of the different LPG qualities within minimum tolerances. The most widely used evaporators are those with single-stage pressure regulation. Figure 4 shows an evaporator with single-stage pressure regulation.

Operating principle
The LPG enters through the inlet connection into the evaporation area. This consists of a labyrinth that is heated by the vehicle's heating circuit to convert the incoming liquid gas into the gaseous state. Pressure of 4 to 15 bar are achieved here. The pressure varies depending on the temperature of the engine coolant and the composition of the LPG (mixture ratio of propane and butane).

The now gaseous medium passes through a transfer passage into the pressure-regulation area, which is closed by a diaphragm. A pre-tensioned spring is seated on the back of the diaphragm. The pre-tension can be adjusted with an adjusting screw in order to effect deviations from the factory output-pressure pre-setting.

The diaphragm is connected via a lever mechanism to a valve on the transfer-passage connection. When the pressure increases, the movement of the diaphragm actuates a control rod, which in turn alters the opening cross-section at the outlet of the pressure-regulation area. Thus the supply is reduced or closed off and no more gas can flow. When the gas pressure decreases in response to increased fuel reduction by the engine, the diaphragm is moved in the opposite direction. The valve is opened via the lever mechanism to allow gas to flow until again an equilibrium of forces acts on the diaphragm and the lever mechanism can close the transfer passage again.

An important characteristic of the quality of the evaporator is the lowest possible fluctuation of the operating pressure in different engine operating states together with the highest possible control precision over the operating time.

In certain countries, owing to contaminants in the LPG, the manufacturer recommends that the diaphragm be replaced after a certain operating time.

Intake-manifold vacuum support
In addition, the response of evaporator pressure regulation is optimized by connecting the intake-manifold pressure to the area above the diaphragm.

Operating values
The inlet pressure is 4 to 15 bar, depending on ambient temperature and gas quality. The output pressure is 0.5 to 3.5 bar, depending on the setting.

Electromagnetic shutoff valve and temperature sensor
The electromagnetic shutoff valve interrupts the gas supply when the engine is not in gas mode. The changeover from gasoline to LPG operation is initiated by the engine ECU as a function of the engine coolant temperature. The minimum coolant-temperature value for changeover is selected to ensure reliable evaporation of the LPG in the evaporator. A temperature sensor is integrated in the evaporator coolant circuit for this purpose. A cold start in gas mode is not possible with these systems.

References
[1] World LPG Gas Association, www.wlpga.org.
[2] www.gas-tankstellen.de.
[3] www.autogastanken.de.
[4] ECE-R67: Standard conditions on the:
I. Approval of specific equipment of motor vehicles using LPG in their propulsion system.
II. Approval of a vehicle fitted with specific equipment for the use of LPG in its propulsion system with regard to the installation of such equipment.
[5] K. Reif: Ottomotor-Management, 4th Ed., Verlag Springer Vieweg, 2015.

Engines fueled by natural gas

Applications

In view of the worldwide efforts to reduce CO_2 emissions (carbon dioxide), natural gas is becoming increasingly important as an alternative fuel. Compressed natural gas (CNG), not to be confused with liquified petroleum gas (LPG), consists primarily of methane. Liquified petroleum gas, on the other hand, consists of propane and butane (see fuels).

Emissions
Compared to gasoline, the combustion of CNG produces approximately 25 % less CO_2. CNG thus generates the lowest local CO_2 emissions of the fossil fuels. Because of the lower CO_2 emissions, the petroleum tax on CNG is reduced in many countries.

Use of biogas
Using biogas (biomethane) can achieve a regenerative fuel supply that further reduces global greenhouse-gas emissions. Biomethane can be produced from plant remains, organic waste and liquid manure, or alternatively from electricity generated from renewable sources. Here, hydrogen is obtained by electrolysis from water and then reacts in a chemical reaction with CO_2 to produce methane. This methane gas obtained from renewable sources is called solar gas, wind gas or simply RE gas (renewable-energy gas).

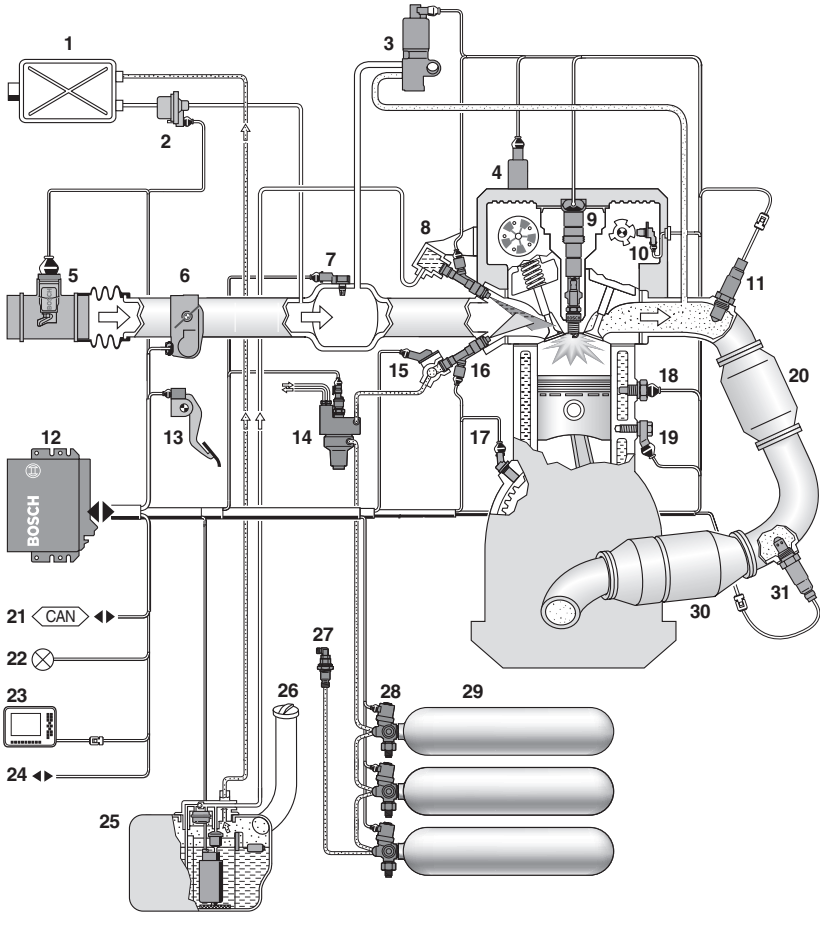
Natural-gas vehicles
Various vehicle manufacturers have in the meantime started to offer natural-gas versions of automobile models from the factory. This ensures that the larger CNG tanks are accommodated more efficiently and conveniently, thereby preventing – as is practically inevitable when these tanks are installed subsequently – the loss of luggage-compartment volume.

The latest information on natural-gas vehicles and the filling-station network in Germany and Europe can be found at the Internet addresses [1] and [2]. Natural-gas vehicles are, for the most part, bifuel vehicles which allow the driver to switch between natural-gas and gasoline operation. There are variants in the form of monovalent vehicles and the concepts referred to as “Monovalent plus”, in which the engine is optimized for natural-gas operation so as to be able to utilize to optimum effect all the favorable properties of natural gas (higher knock resistance, lower CO_2 and pollutant emissions). “Monovalent plus” vehicles still feature a small gasoline tank (less than 15 *l*) so that they can continue to be driven in gasoline mode in the event of there being no natural-gas filling station nearby.

Storage of natural gas
Natural gas can be stored in compressed form at a pressure of 200 bar as CNG (compressed natural gas) or in liquid form at –162 °C as LNG (liquified natural gas). In view of the great expense involved in storing the gas in liquid form, storing the gas in compressed form at 200 bar has become the standard method. In spite of being stored at high-pressure, natural gas has a lower energy-storage density than gasoline such that a tank with four times the volume is required for the same energy content.

Figure 1: Spark-ignition engine with natural-gas or gasoline operation

1 Carbon canister, 2 Canister-purge valve, 3 Exhaust-gas recirculation valve,
4 Camshaft adjuster, 5 Hot-film air-mass meter,
6 Electronically actuated throttle valve (ETC),
7 Intake-manifold pressure sensor, 8 Gasoline rail with gasoline injectors,
9 Ignition coil with spark plug, 10 Camshaft sensor,
11 λ sensor before catalytic converter,
12 Bifuel Motronic ECU, 13 Accelerator-pedal module,
14 Natural-gas pressure-control module with integrated gas shutoff valve and high-pressure sensor,
15 Gas rail with natural-gas pressure and temperature sensor, 16 Natural-gas injector,
17 Crankshaft speed sensor, 18 Engine-temperature sensor, 19 Knock sensor,
20 Primary catalytic converter, 21 CAN interface,
22 Engine warning light, 23 Diagnosis interface, 24 Interface to immobilizer ECU,
25 Gasoline tank with integrated electric fuel pump, 26 Gasoline filler neck,
27 Natural-gas filler neck, 28 High-pressure shutoff valve on natural-gas tank,
29 Natural-gas tank, 30 Main catalytic converter,
31 λ sensor after primary catalytic converter.



Design

Components
Natural-gas automobiles are equipped almost exclusively with spark-ignition engines. They are extended to include the following natural-gas components (Figure 1, [3]):
– Natural-gas filler neck,
– Natural-gas tanks,
– High-pressure shutoff valves on the natural-gas tanks,
– Natural-gas pressure-control module with integrated gas shutoff valve and high-pressure sensor,
– Gas rail with natural-gas injectors and combined natural-gas low-pressure and temperature sensor.

Operating principle
The air drawn in by the engine is routed via the air-mass meter and the electronically actuated throttle valve to the intake manifold. From there, it enters the combustion chamber through the intake valves (Figure 1). The gas stored in the natural-gas tank at 200 bar flows through a high-pressure shutoff valve on the tank into the natural-gas pressure-control module, which lowers the gas pressure to a constant operating pressure of approx. 5 to 10 bar (absolute). It then flows through a flexible low-pressure line into a common gas rail that supplies the natural-gas injectors.

Engine management for natural-gas vehicles
Both double-ECU concepts (one ECU for gasoline operation and one for natural-gas operation) and single-ECU concepts (gasoline and gas functions integrated into a single ECU) are used. In some bifuel vehicles, the driver can use a switch to select between gas and gasoline operation. In most models, the switchover occurs automatically and the vehicle is driven in natural-gas mode until the tank contents are used up. Then the vehicle is switched over automatically to gasoline mode.

The high-temperature sensor mounted on the pressure-control module supplies the engine-management system with information on the current fill level of the natural-gas tank and also facilitates leakage diagnostics. The combined natural-gas low-pressure and temperature sensor fitted on the gas rail enables the engine-management system to correct the valve injection timing in such a way that a stoichiometric mixture can be pilot-controlled in the intake manifold even if the gas density fluctuates. The engine-management system uses an adaptation algorithm to adapt to changing gas qualities.

The other sensors and actuators of the engine-management system are mostly identical to those in a gasoline engine.

Mixture formation
Virtually exclusively bifuel and "Monovalent plus" vehicles are available today. These vehicles can operate both on natural gas and on gasoline. Either manifold injection or gasoline direct injection may be used to inject the gasoline.

Natural-gas manifold injection
In most natural-gas engines, the gas is injected, as with gasoline injection, into the intake manifold. The low-pressure gas rail supplies the injectors, which inject natural gas at intermittent intervals into the intake manifold. Mixture formation is improved in comparison with gasoline injection, as natural gas, because of the completely gaseous supply of fuel, does not condense on the intake-manifold walls and does not deposit a film there. This has a positive effect on emissions, particularly during the warm-up phase.

Non-supercharged engines that are fueled by natural gas generally have a 10 to 15 % lower output. This is due to the lower volumetric efficiency attributed to displacement of the inducted air caused by the injected natural gas. The vehicle engines can however be optimized specially for natural gas as the fuel. The extremely high knock resistance of natural gas (up to 130 RON) provides for higher engine compression and is ideally suited to supercharging. In this way, the volumetric-efficiency losses can even be overcompensated. Combined with reduced piston displacement (downsizing), efficiency increases due to additional dethrottling and reduced friction.

Natural-gas direct injection
Natural gas can also be injected directly into the combustion chamber. In this way, the volumetric-efficiency losses can be completely avoided and in turbocharged engines the supercharging efficiency, especially at low speeds, can be improved by scavenging. Scavenging involves engine operation with extended overlapping of the intake- and exhaust-valve opening times. The drawbacks of natural-gas direct injection are the greater complexity of the injectors (tightness, temperature stability) and the fact that, due to the necessary higher natural-gas injection pressure dictated by the concept, the tank cannot be drained as far as in engines with natural-gas manifold injection.

Exhaust emissions
Compared with gasoline engines, natural-gas vehicles are characterized by approximately 25 % lower CO_2 emissions. The reason for this is the better hydrogen/carbon ratio (H/C ratio) of almost 4:1 (gasoline: approximately 2:1). This results in the creation of more water and less CO_2 during combustion.

Apart from the virtually particulate-free combustion, in conjunction with a three-way closed-loop catalytic converter, only very low levels of the pollutants NO_x, CO and HC are produced. As a rule, the catalytic converter has a higher precious-metal load than is the case in gasoline engines, enabling it to better convert the HC emissions in the exhaust gas that mainly consist of chemically stable methane and to counteract the higher "light-off temperature" (minimum temperature of the catalytic converter for conversion) for natural gas. Methane is, however, classified as nontoxic.

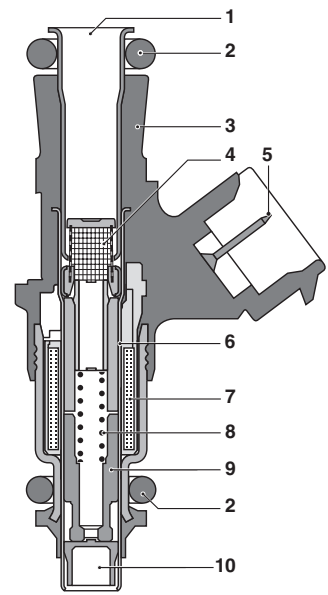
Natural-gas vehicles comply with the current exhaust-emission limits, while trucks and natural-gas buses also comply with the EEV limits (Enhanced Environmentally friendly Vehicle, [4]). The natural-gas engine offers distinct advantages over gasoline and diesel engines, also in particular with regard to the non-limited pollutant emissions.

Components

Natural-gas injector for manifold injection
Function
In order to supply the internal-combustion engine with gaseous fuel, it is necessary to meter through the natural-gas injectors a much greater volume of gas than gasoline in a conventional gasoline engine. This requirement places specific demands on the design of the natural-gas injector, which must be adapted to the greater gas volume in its cross-sections. Even the high flow velocities that are encountered call for a special form of flow routing in order to reduce pressure losses in the injector.

In highly supercharged engines, the intake-manifold pressure can rise up to 2.5 bar (absolute). In order to suppress the influence of the intake-manifold pres-

Figure 2: Natural-gas injector (Bosch)
1 Pneumatic port, 2 Sealing ring, 3 Valve housing, 4 Filter strainer, 5 Electrical connection, 6 Sleeve, 7 Solenoid coil, 8 Valve spring, 9 Solenoid armature with elastomer seal, 10 Valve seat.



sure on the mass flow, it is necessary for the pressure ahead of the nozzle at the narrowest cross-section (throttling point) accepted as the nozzle above critical to be at least twice as high as the maximum intake-manifold pressure (pressure after the nozzle). The gas then flows at the speed of sound, irrespective of the absolute pressure after the nozzle. The mass flow is therefore not influenced by the variable intake-manifold pressure. While possible pressure losses ahead of the throttling point are taken into account, a minimum operating pressure of 7 bar (absolute) has proven to be advantageous.

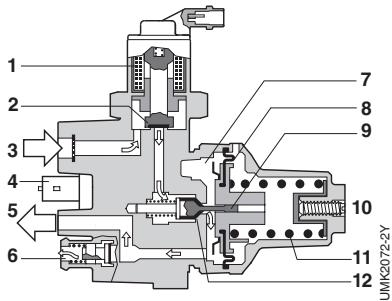
Design and operating principle
The solenoid armature (Figure 2) is guided in a sleeve. The armature has fuel flowing through it on the inside and has an elastomer seal at the discharge end. This seal closes on the valve seat and thereby seals off the fuel supply from the intake manifold. When energized, the solenoid coil effects the necessary force to lift the solenoid armature and open the metering cross-section (throttling point in the valve seat). When the coil is at zero current, the valve spring holds the injector closed.

Flow-optimized geometry
Thanks to design measures in the routing of the flow, the pressure loss is minimized ahead of the throttling point in order to ensure the greatest possible mass flow. Furthermore, the narrowest cross-section and thus the throttling point is located at the discharge end after the seal. The speed of sound prevails here such that the valve represents approximately an ideal nozzle in physical terms.

Sealing geometry
The natural-gas injector is fitted with an elastomer seal and is similar in terms of its seal-seat geometry to shutoff valves for pneumatic applications. The elastomer increases the tightness of the seal against metallic needle valves.

Damping in the elastomer also prevents "rebounding", i.e. a repeated, unwanted opening of the solenoid armature during the closing operation, and thus increases metering precision.

Figure 3: Pressure-control module
1 Solenoid coil of electromagnetic gas shutoff valve,
2 Seal,
3 High-pressure inlet with sinter filter,
4 High-pressure sensor,
5 Low-pressure outlet,
6 Safety pressure-relief valve,
7 Low-pressure chamber,
8 Pressure diaphragm with mounting,
9 Control rod,
10 Adjusting screw,
11 Spring,
12 Control plunger and throttling orifice.



Pressure-control module
Function
The function of the pressure-control module is to reduce the pressure of the natural gas from tank pressure (up to 200 bar) to the nominal operating pressure (typically 5 to 10 bar). At the same time, the operating pressure should be kept constant within specific tolerances through all operating states or, depending on the engine's operating point, regulated to a specified pressure. The typical operating pressure of present-day system is usually about 5 to 10 bar (absolute). There are also systems which operate at pressures starting from 2 bar ranging up to 11 bar.

Design
Mechanical version:
Today, mainly mechanical diaphragm- or plunger-type pressure regulators are used. Pressure reduction is effected by means of throttle action and can occur either in one single stage or in several stages.

Figure 3 shows the sectional view of a single-stage diaphragm-type pressure regulator. A sinter filter (pore size approx. 40 µm), a gas shutoff valve, and a high-pressure sensor are integrated into the pressure-control module on the high-pressure side. The sinter filter is designed to retain solid particulates in the gas flow. The gas shutoff valve enables the gas flow to be interrupted, for example when the vehicle is stopped. The high-pressure sensor is used to determine the natural-gas supply in the tank and for diagnostics.

A pressure-relief valve is mounted on the pressure regulator on the low-pressure side. In the event of a fault in the pressure regulator, this valve prevents damage to components in the low-pressure system.

In normal control operation, the gas cools down very dramatically as it expands as a result of the Joule-Thompson effect. In order to prevent freezing, the pressure-control module has an integrated heater, which is connected to the vehicle's heating circuit (not shown in the figure).

The controlled operating pressure is dependent on the diaphragm size and the initial spring tension. To adjust the pressure level, the initial spring tension is set and sealed by means of an adjusting screw in the factory.

Electromechanical version:
As well as purely mechanical pressure-control modules there are also electromechanical versions. These typically consist of an initial mechanical pressure-control stage, which reduces the tank pressure to a mean pressure of approximately 20 bar. In a second stage the pressure is further reduced by an electromagnetically actuated control valve to the operated pressure specified by the engine-management system, and electronically regulated. The advantages of electromechanical pressure regulators are the increased pressure-control precision and the variability of the control pressure. In this way, the injection pressure can be decreased and thus the metering precision of the natural-gas injectors increased in engine low-load phases, e.g. at idle.

Operating principle of mechanical pressure-control module
The gas flows from the high-pressure side through a variable throttling orifice into the low-pressure chamber, where the diaphragm is situated. The diaphragm controls the opening cross-section of the throttle via a control rod. When the pressure in the low-pressure chamber is low, the diaphragm is forced by the spring in the direction of the throttling orifice, which opens to allow the pressure to increase on the low-pressure side. In the event of excessive pressure in the low-pressure chamber, the spring is compressed more sharply and the throttling orifice closes. The decreasing cross-section of the throttling orifice reduces the pressure on the low-pressure side. In stationary operation, the throttling cross-section required for the operating pressure is set, while the pressure in the low-pressure chamber remains extensively constant.

Minimal operating-pressure fluctuations in the event of load changes testify to the quality of the pressure regulator.

References
[1] www.erdgas.info/erdgas-mobil/.
[2] www.gibgas.de.
[3] T. Allgeier, J. Förster: 2nd International CTI Forum, Stuttgart, March 2007; Einspritzsysteme – Motormanagementsystem und Komponenten für Erdgasfahrzeuge.
[4] www.bmub.bund.de/themen/luft-laerm-verkehr/luftreinhaltung/eev-standard/.

Engines fueled by alcohol

Applications

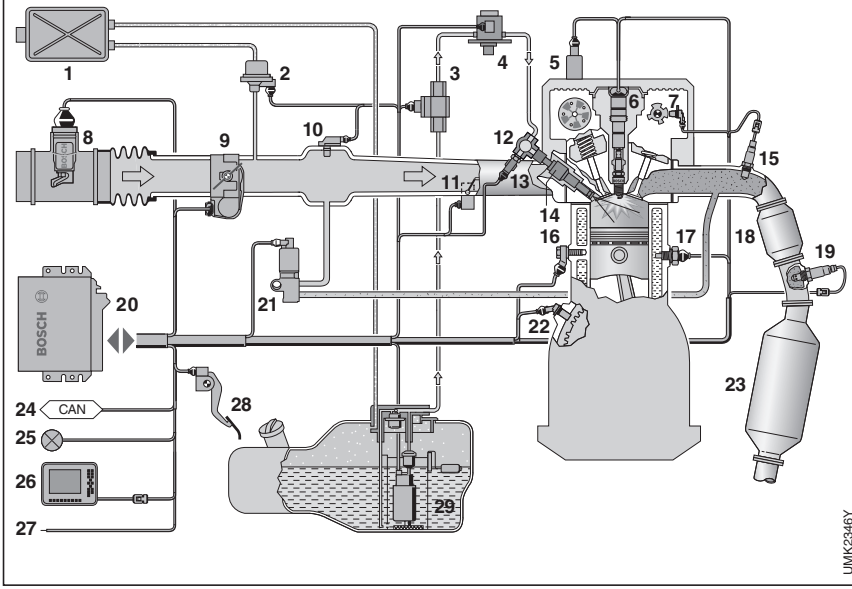
Motivation

Ever-shrinking fossil-fuel resources and the heightened demand for reduced carbon-dioxide emissions necessitate a rethink of tried-and-tested drive concepts. Rising crude-oil prices make alternative energies increasingly economical and there is the increasing public perception that sustained mobility also requires a sustained energy basis.

In Brazil, where ethanol has been used as a fuel since the 1980s and is available as E 24 (ethanol/gasoline mixture with 24 % ethanol content) and E 100 (pure ethanol) at filling stations (Table 1), approximately 50 % of fuel demand is covered. Alongside this, the USA and Europe have in the last few years increased their activities to promote E 85 as an alternative fuel type. Thus, aside from the potential to reduce CO_2, the focus is also aimed at securing reliability of supply and making fuel

Figure 1: Gasoline engine with gasoline direct injection for flex-fuel operation

1 Carbon canister, 2 Tank vent valve, 3 Ethanol sensor (optional),
4 High-pressure pump, 5 Camshaft adjuster, 6 Ignition coil with connected spark plug,
7 Camshaft phase sensor, 8 Air-mass meter, 9 Throttle device (ETC),
10 Intake-manifold pressure sensor, 11 Swirl control valve,
12 Fuel rail, 13 High-pressure sensor, 14 High-pressure injector,
15 λ sensor, 16 Knock sensor, 17 Engine-temperature sensor, 18 Primary catalytic converter,
19 λ sensor, 20 Engine ECU, 21 Exhaust-gas recirculation valve,
22 Engine-speed sensor, 23 Main catalytic converter, 24 CAN interface,
25 Malfunction indicator lamp, 26 Diagnosis interface, 27 Interface to immobilizer,
28 Accelerator-pedal module, 29 Fuel-supply module with low-pressure pump.



supply independent of fossil sources and imports.

System description
Flex-fuel systems are engine-management and component systems which can be operated with any mixture ratios of gasoline and ethanol between two limit values. These mixture ratios are created in the fuel tank by pumping in fuels with varying ethanol content. On the one hand, the capacity of a system to detect varying ethanol content in the fuel and to adapt its behavior (e.g., injection and ignition) accordingly plays a role. On the other hand, the question of component suitability is very important in making safe, robust operation possible over the entire service life demanded and expected by vehicle users.

Figure 1 shows the system diagram of a gasoline engine for flex-fuel operation. It is to a large extent the same as a conventional gasoline engine with gasoline direct injection.

Ethanol as an alternative fuel

Sustainability
The potential to reduce CO_2 emissions is established in part by the more favorable ratio of carbon to hydrogen content of alcohol fuels, but primarily by the positive CO_2 balance in the course of extraction from renewable raw materials. It is however important here to consider the source of the produced alcohol fuel. Thus, for example, the extraction of methanol from coal (by means of the CTL process, Coal to Liquid) has an extremely poor CO_2 balance, whereas the extraction of ethanol from parts of plants has a positive balance.

Markets for flex-fuel vehicles
Today, the admixture of methanol in appreciable quantities plays, on account of local factors, a role only in China. Moreover, methanol, because it is toxic and harmful to health, must be viewed as highly critical. In the following, flex-fuel operation refers to operation with ethanol fuels.

Table 1: Composition of ethanol fuels

Fuel	E0…E5	E 10	E 24	E 85	E 100
Maximum ethanol content	5 %	10 %	24 %	85 %	93 %
Maximum water content	<1 %	<1 %	1 %	1 %	7 %
Minimum gasoline content	95 %	90 %	76 %	15 %	0 %
Relative energy content referred to gasoline	100 %	97 %	91 %	70 %	61 %
Regional distribution	Europe	USA, EU	Brazil	USA, Sweden, partly EU	Brazil

Table 2: Properties of ethanol

Characteristic	Unit	Gasoline	Ethanol	Effects
Energy density (per volume)	kJ//	32,500	21,200	– greater injected fuel quantity required
Energy density (per mass)	kJ/kg	43,900	26,800	– critical for cold start
Air/fuel ratio	–	14.8	9.0	– more wall film + more power, better efficiency (greater air cooling)
Boiling point	°C	25 to 215	78	
Enthalpy of evaporation	kJ/kg	380 to 500	904	
Octane number	RON	>91	108	+ higher knock limit + more power, higher efficiency
H/C ratio	–	2.3	3	

Flex-fuel refers to mixtures of gasoline and ethanol in any variable ratio. Today, a distinction is drawn primarily between two mixture ranges of ethanol mixtures in the flex-fuel sector: Mixtures with mixture ratios between pure gasoline and E 85 (in Europe particularly in Sweden and in North America) and between E 24 and E 100 in Brazil. The lower limit towards E 18 is currently being extended in Brazil.

Another specific feature is that E 100 fuel according to the Brazilian standard (Resolution ANP #19 – Technical Regulation, [1]) may contain up to 7.5 % water (EHC, Hydrated Ethanol Fuel), but in real terms in the field contains 10 to 15 % water. Also associated with this increased water content is increased conductivity and the increased presence of salts. This is the reason why there are two separate variants of a component for the E 85 and E 100 application cases and also specific adaptations are made in the electronic engine management. It is important here than an E 85 system is not suitable for E 100.

Essentially, flex-fuel systems are implemented with manifold and direct injection. In the case of direct injection it is possible in particular to use the high knock resistance of ethanol to good effect, which is advantageous with regard to the full-load quantity demand due to less fuel enrichment.

Properties of ethanol and resulting requirements

The specific properties of ethanol that deviate from regular gasoline are set out in Table 2. These impact on system performance and on the components in the system that come directly or indirectly into contact with fuel.

Important factors are the lower energy density (greater quantity demand per unit of air) and the deviating stoichiometry of ethanol caused by the chemical composition (oxygen is already contained in part), which results in dramatically increased fuel-system delivery and injected fuel quantities. Critical to the starting response at very low temperatures are the high boiling temperature of 78 °C and the high heat of evaporation compared with the low-boiling fractions of gasoline. However, for operation at high engine temperatures and at the knock limit, advantageous effects are also to be expected due to the high heat of evaporation and better knock resistance.

These mentioned properties of ethanol and their effects make it necessary to adapt the internal-combustion engine itself, but also the engine-management system and its components accordingly. It is also important that the ethanol fuels not be present as pure fuels with a fixed (and thus known) ethanol content in the fuel tank, but that the mixing of different ethanol fuels lead to any mixture ratios. The mixtures can moreover also be supplied by regionally and seasonally deviating ethanol admixtures by the filling stations and in the USA by “blender pumps” in the system. Thus, a reliable and accurate “on-board recognition” of the current ethanol content is required.

The higher corrosivity of ethanol must be taken into account for the components installed in a system. To achieve evaporation limit values the higher permeation tendency of ethanol molecules through pipe and component materials plays a significant role.

Flex-fuel concepts for the markets

Vehicle manufacturers pursue different engine concepts and marketing strategies for the regions. With regard to the degree of adaptation of the systems there are therefore different aspects which conform to the aimed-for strategy.

Markets

USA market

In the US market as a matter of priority only those adaptations which are necessary to be able to offer the vehicles as flex-fuel vehicles are carried out. This involves the fundamental compatibility for flex-fuel without specifically exploiting the advantageous ethanol properties. In real vehicle operation it is used only occasionally with higher and variable ethanol contents.

European market

In contrast, a different concept is pursued in Europe whereby both the increased efficiency of the engine and the enhanced performance in E 85 operation are exploited as advantages, and are emphasized in the marketing of the vehicle.

Brazilian market

Only in Brazil is it possible up to now to observe the last logically consistent step where the engine manufacturers also increase the compression ratios of the engines since in this market the fuels offered already generally contain at least 24 % ethanol and consequently a higher octane number (and thus a higher knock resistance) is obtained.

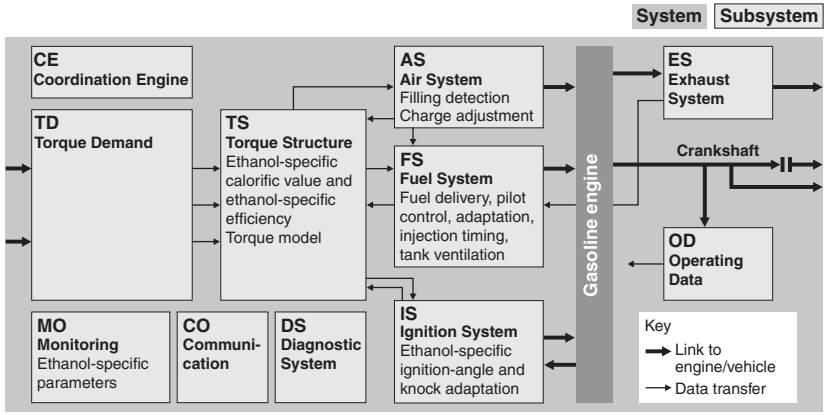
System adaptations

Extensive adaptations are required in the management system of the engine ECU and its functions. Specifically, the following engine-management functions are affected (Figure 2):

- Air system: e.g. filling detection and cylinder-charge control,
- Fuel system: e.g. quantity pilot control and injection timing, mixture adaptation,
- Exhaust system: e.g. enabling of mixture control,
- Ignition system: e.g. moments of ignition, ignition energy,
- Torque patterns: e.g. deviating efficiency with regard to λ and ignition,
- Diagnostic system and monitoring; e.g. fuel-quality detection, diagnostic thresholds.

When it comes to system adaptation the deviating properties of ethanol with regard to gasoline must be taken into account in order to avoid diminished system performance or to achieve improved perfor-

Figure 2: Software structure of engine management for flex-fuel operation



mance in partial ranges. The properties listed in Table 3 can be used to good effect, while those listed in Table 4 necessarily require an adaptation.

In part, however, new developments of functionalities are required. The following are mentioned here by way of example:
– Special cold-start strategies for achieving emission limits at lower temperatures (e.g. –7 °C) and for fundamentally ensuring engine starting at very low temperatures (–30 °C).
– Detection of a more pronounced impingement of fuel in the engine oil and interception of increase degassing during warm-up.
– Detection of alternating ethanol contents in the fuel tank and diagnosis of ethanol detection (sensor-based or sensorless).

Table 3: Advantageous use of ethanol properties (with regard to gasoline)

Properties of ethanol	Advantageous use
Higher knock resistance	Optimum ignition, better efficiency, more power
Higher enthalpy of evaporation	Better charge
Faster combustion	Optimum maximum combustion
Lower combustion temperatures	Reduced component protection with lower consumption

Flex-fuel components

Flex-fuel systems are virtually identical in design to gasoline injection systems; however, the demands imposed on the components differ. With regard to component suitability, it is necessary to consider both the fuel-carrying components and those components that potentially come into contact with fuel vapors or combustion gases of ethanol. An overview of the components to be adapted in a flex-fuel system is provided in Table 5.

Changes in the components must be made with regard to the following influences:
– corrosion of metallic materials,
– swelling or embrittlement of plastic and rubber parts,
– functional capability of gaskets,
– increased permeation of fuel molecules through fuel-carrying components.

This is done specifically by means of the following measures:
– use of special alloys (e.g. stainless steels),
– surface coating of metallic materials, particularly in mechanically stressed contact pairs (above all on aluminum materials, e.g., Nikasil coatings of nickel and silicon carbide for aluminum engine blocks),
– use of adapted plastic and rubber compounds,

Table 4: Disadvantageous consequences of ethanol properties (with regard to gasoline)

Properties of ethanol	Disadvantageous consequences
Higher enthalpy of evaporation	Increased requirements imposed on cold start below 10 °C, larger wall film in cold operation
Higher stoichiometric consumption	Greater required injected fuel quantities, longer injection periods, repercussions for the combustion process, greater fuel impingement in the engine oil during cold start
Changed H/C ratio	More water impingement, later dew-point end of λ sensor

Table 5: Overview of ethanol-specific components

Component	Adaptation
Fuel-supply module	Fuel pump for E 85 or E 100, insulated electrical connections, Fuel-level sensor ethanol-resistant
Pressure regulator	Ethanol-resistant version
Fuel rail	Ethanol-resistant version
Intake-manifold fuel injector	Extended metering range
High-pressure injector	Stainless-steel versions, adapted metering range
High-pressure pump	Stainless-steel versions, adapted metering range

– selection of specific gasket and filter materials,
– design adaptations,
– electric insulation of electrical contacts in the fuel tank (electric fuel pump, fuel-level sensor) for E 100 due to the water content of up to 15 %.

Increased requirements with regard to greater flow rates must be taken into account for specific components. This is particularly important in relation to the fuel injectors since the maximum flow rate is necessary for the increased demand in ethanol operation, but still the minimum metering quantity occurs in gasoline operation. This adaptation is effected by the selection of a greater flow rate from the range of variants. The unchanged minimum quantity is implemented by a shortened activation duration.

References

[1] http://www.itecref.com/pdf/Brazilian_ANP_Fuel_Ethanol.pdf.
[2] A. Böge, W. Böge: Handbuch Maschinenbau, 23rd Ed., Verlag Springer Vieweg, 2016.
[3] K.-H. Grote, B. Bender, D. Göhlich (Editors): Dubbel: Taschenbuch für den Maschinenbau, 25th Ed., Verlag Springer Vieweg, 2018.
[4] GVR Gas Vehicle Report, June 2012.

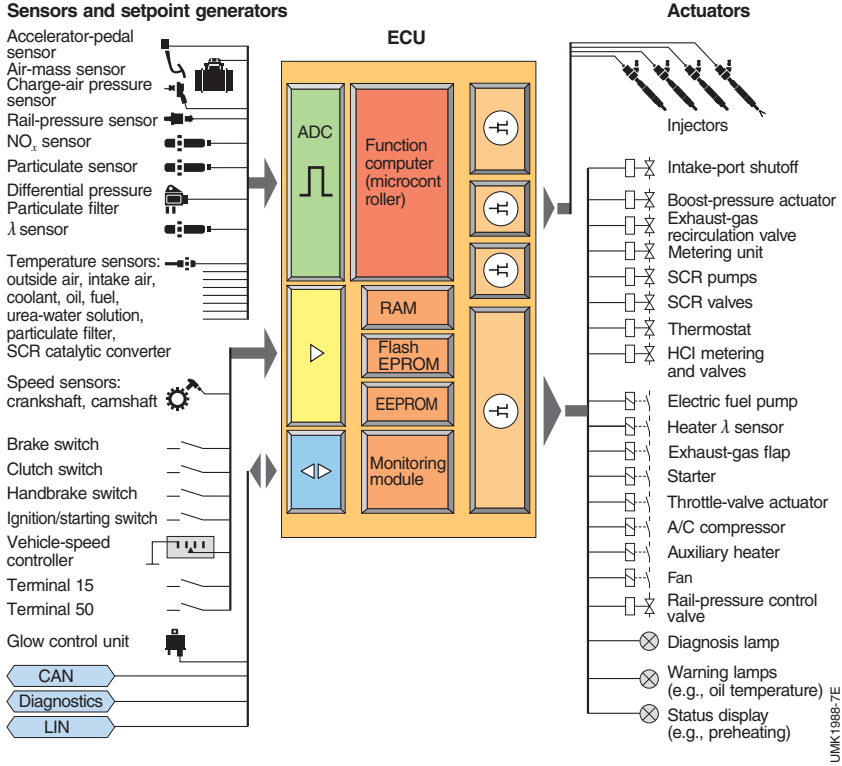
Electronic Diesel Control

Description of the engine-management system

The diesel combustion process requires on principle an injection of fuel into the combustion chamber. Modern diesel engines operate with direct injection; the cylinders have a common high-pressure pump and a high-pressure accumulator constantly under high pressure (up to 3000 bar) that is common to all the cylinders (common rail). Injection is initiated by the opening of the fuel injectors, which are activated electronically. Electronic engine management is an absolute precondition for the common-rail injection process.

Furthermore, the statutory requirements with regard to emissions and customer expectations in relation to a modern vehicle can only be fulfilled with an electronic engine-management system. The open- and closed-loop control system – both for the earlier injection systems and for the current common-rail system – is referred to by Bosch as Electronic Diesel Control (EDC).

Figure 1: Components for Electronic Diesel Control (EDC) of a common-rail system



Requirements

The primary and obvious purpose of the engine-management system is to deliver the desired power (i.e., torque times rotational speed) to the diesel engine's output shaft. Further requirements include facilitating the best possible engine operation with regard to fuel consumption (and consequently CO_2 emissions), exhaust-gas emissions, power development and driving comfort, and diagnosing the status of the ECU, the sensors, and the actuators. Furthermore, the engine-management system must protect the engine and the drivetrain against unacceptable operating conditions and overloading.

To achieve this, many operating parameters must be acquired with sensors and processed with algorithms – these are computing processes that take place according to an established scheme. The results are signal curves for activating the actuators. Electronic Diesel Control (EDC) refers to the full range of sensors, engine management by engine control unit, and actuators (Figure 1). Sometimes the engine control unit in itself is referred to as EDC.

Requirements with regard to function groups
The requirements for the individual engine function groups can be derived from the requirements with regard to the engine.

Detection of driver input
– Accelerator-pedal sensors (the accelerator-pedal position is detected for safety reasons with redundant sensors).

Requirements with regard to the injection system
– High injection pressures for atomizing the fuel and reducing the raw particulate emissions,
– rate-of-discharge curve control (multiple injection, pilot injection, and secondary injection) to improve combustion, emissions, and noise,
– variation of injected fuel quantity and start of injection to suit every operating condition,
– tight tolerances for injection time and injected fuel quantity and maintenance of high precision over the service life of the system.

Requirements with regard to the air system
– Single- or multi-stage supercharging with intercooling (charge-air cooling),
– controlled exhaust-gas recirculation with cooling to reduce NO_x emissions,
– if necessary additional low-pressure exhaust-gas recirculation (EGR) to improve the cylinder charge compared with pure high-pressure EGR.

Requirements with regard to the exhaust system
– Exhaust-gas treatment with oxidation-type catalytic converter, particulate filter, NO_x storage catalytic converter, and selective catalytic reduction (SCR) or both,
– regeneration of the catalytic converters,
– rapid heating of the catalytic converters after engine start.

Requirements with regard to CO_2 emissions
– Operation of the engine in the optimum program map,
– start/stop functions,
– thermal management,
– consumption-optimized auxiliary systems, e.g., speed-controlled water pump or electric fans,
– controlled generator operation (for recuperation in coasting/overrun mode).

Other requirements
– Rapid engine start, irrespective of the ambient conditions,
– idle-speed control to load-independent setpoint value,
– comfort and convenience functions such as cruise control, speed limiting,
– comprehensive diagnostics in real time.

Operating principle and architecture

Modern vehicles have a multitude of (up to 100) ECUs which are connected to each other via communication buses (usually CAN buses). Engine management is integrated in the architecture of the overall vehicle system so that data from other electronic systems such as for example antilock braking system (ABS), transmission-shift control or driving-dynamics control are taken into consideration in the open- and closed-loop control of the diesel engine. These can be control interventions on the torque level (e.g., engine-torque reduction when engaging the automatic transmission, adapting the engine torque to the slip of the wheels) or even status information (e.g., injection enable by the vehicle immobilizer).

A number of tasks can also be combined in an ECU, depending on the strategy and philosophy of the vehicle manufacturer. The number of sensors and actuators connected to the engine control unit is dependent on the architecture of electronic systems (see Architecture of electronic systems). The number of functions can be within wide limits, i.e., between the extremes "injection control unit" on the one side and "engine management with integrated vehicle control and exhaust-gas treatment" on the other. Figure 2 shows examples of executed systems.

All influencing variables have an impact on the amount of fuel to be injected and the start of injection. They therefore leave their mark on predefined algorithms with which the corresponding control variables are calculated in the ECU and from them the activation signals for the actuators are generated. In the process no states that lead to a vehicle operating condition that is not wanted by the driver may occur during vehicle operation. This is made possible by a comprehensive monitoring concept which must detect occurring deviations in real time and with regard to possible effect introduce appropriate measures (e.g., torque limitation or emergency operation in the idle-speed range). To this end the ECU contains a monitoring module in addition to the main computer (see Electronic control unit (ECU)).

System blocks

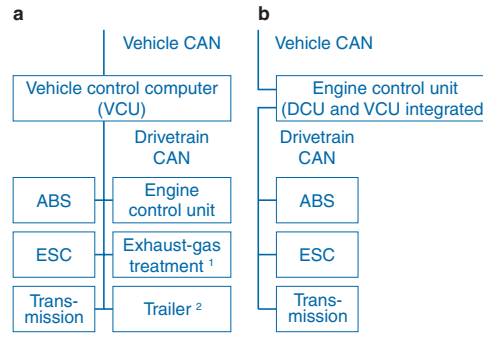
The complete Electronic Diesel Control (EDC) system – consisting of sensors and setpoint generators, the ECU, and the associated actuators (Figure 1) – is subdivided into different blocks, which in each case form a functional or structural unit. These blocks are called system blocks.

Sensors

The electronic engine control unit uses sensors to acquire the operating data required for open- and closed-loop engine control which are converted into electrical signals so that they can be processed in

Figure 2: Vehicle architecture

a) Architecture with vehicle control computer and exhaust-gas treatment ECU,
b) Vehicle functions and exhaust-gas treatment integrated in the engine control unit.
VCU Vehicle Control Unit (vehicle control computer),
DCU Dosing Control Unit (dosing of reducing agent for exhaust-gas treatment with SCR),
ABS Antilock Braking System,
ESC Electronic Stability Control (Electronic Stability Program, driving-dynamics control).
[1] SCR systems with DCU,
[2] Commercial vehicles only.



the “ECU” system block. Sensors detect physical and chemical quantities and thereby provide information about the engine's current operating condition. Examples of such sensors are:

Driver-input detection:
– accelerator-pedal position,
– brake switch.

Vehicle-condition detection:
– driving speed,
– clutch switch,
– brake pressure.

Ambient conditions:
– air temperature,
– air humidity,
– atmospheric pressure.

Engine-condition detection:
– speed sensor for detecting the crankshaft position and calculating the engine speed,
– phase sensor for detecting the camshaft position,
– intake-air temperature,
– air mass,
– boost (charge-air) pressure,
– charge-air temperature,
– coolant temperature,
– excess-air factor (via λ sensor).

Condition, exhaust-gas treatment (integrated systems):
– temperature, pressure, and tank fill level of urea/water solution,
– catalytic-converter temperatures,
– particulate-filter differential pressure.

Electronic control unit (ECU)
The signals of the sensors and setpoint generators are processed in the ECU according to predefined mathematical computing rules (control algorithms). Input variables (e.g., fuel quantity) are calculated via the algorithms and parameters stored in the ECU and to this end the electrical output signals are generated for the respective actuators (e.g., solenoid valve in the injector). The ECU also serves as an interface to other electronic systems (e.g., antilock braking system (ABS) and diagnostics).

Actuators
The third system block comprises the actuators, which convert the electrical output signals from the ECU into mechanical quantities. Examples of actuators:

Driver information:
– warning lamps (e.g., malfunction indicator lamp),
– status displays.

The discrete lights and displays are increasingly being replaced by electronic displays, which receive their information via a data bus.

Diesel injection:
– fuel injectors,
– metering unit,
– pressure-control valve.

Air system:
– EGR valve (exhaust-gas recirculation),
– EGR-cooler bypass,
– charge-pressure actuator for VTG (variable turbine geometry),
– λ sensor heating,
– exhaust-gas flap.

Exhaust-gas treatment:
– SCR heater for tank and lines,
– SCR pumps,
– SCR valves,
– HCI metering and valves (Hydro Carbon Injection, diesel injection into the exhaust-gas system).

Engine cooling and engine protection:
– fan,
– water pump,
– oil pump,
– thermostat.

Other:
– starter,
– engine brake.

Data processing

Operating principle
The main function of Electronic Diesel Control (EDC) is to control the injected fuel quantity, the start of injection, and the injection pressure. The engine control unit activates various actuators in all systems. For all components to operate efficiently, the EDC functions must be precisely matched to every vehicle and every engine (through application and calibration). This is the only way to optimize component interaction.

The ECU analyzes the incoming signals from the sensors and limits them to the permissible voltage level. Some input signals are also checked for plausibility. Using these input signals together with stored program maps, the microcontroller in the ECU calculates injection timing and its duration. This information is then converted to a signal characteristic which is aligned to current crankshaft and camshaft positions. The computing program is part of the ECU software, which also comprises further functions. The required degree of accuracy together with the diesel engine's outstanding dynamic response requires high-level computing power.

The output signals are applied to output stages, which provide adequate power for the actuators (e.g., high-pressure solenoid valves for injection, EGR positioner, and charge-pressure actuator). Apart from this, a number of other auxiliary-function components (e.g., glow relay and air-conditioner compressor) are triggered. Faulty signal characteristics are detected by output-stage diagnosis functions for the solenoid valves.

Furthermore, signals are exchanged with other systems in the vehicle via the interfaces. The engine control unit monitors the complete fuel-injection system as part of a safety strategy.

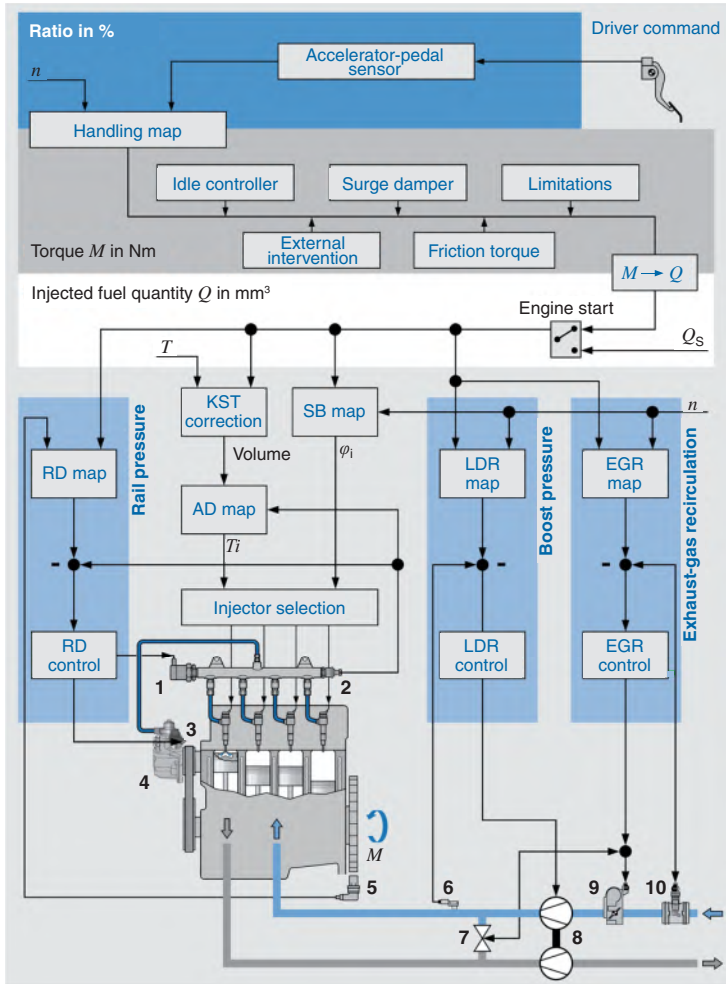
Torque structure in the EDC
Historically, quantities were used to make calculations in the engine control unit; this approach has in the meantime become out-dated. All modern diesel-control systems operate with torque control because torque is the dominant physical variable in the drivetrain. All the power demands on the engine are coordinated and converted into a torque request. In the torque coordinator these demands are prioritized by internal and external consumers as well as further inputs with regard to engine efficiency. The driver when accelerating directly requests a torque to be set via the accelerator pedal (with an accelerator-pedal sensor) (Figure 3). Irrespective of this, other external vehicle systems request via the interfaces a torque which is obtained from the power demand of the components (e.g., air conditioner, alternator). The engine management calculates from this the resulting engine torque and triggers the actuators of the injection and air systems. This has the following advantages:

– No system has a direct influence on the individual engine-management functions (boost pressure, injection, preheating). The engine management can thus for the external requirements take into consideration higher-level optimization criteria (e.g., exhaust-gas emissions, fuel consumption) and then activate the engine in the best possible way.
– Many functions which do not directly concern the management of the engine can run uniformly for diesel- and gasoline-engine management systems. This facilitates the exchange of software.
– Expansions of the system can be quickly implemented.

The definitive sensor for determining the driver command is the accelerator pedal (accelerator-pedal position in %). The accelerator-pedal position is tracked into the driver-command torque (in Nm) via the handling map. In the subsequent "torque path" the influences of idle controller, governor, external interventions (e.g., slip-control torque reductions or torque increases), limitations, etc. are calculated in addition to the driver-command torque. The calculations are made depending on the type on the clutch-, transmission-, or engine-torque level. The final result is the command torque at the engine shaft. In conclusion, the requested internal engine torque is obtained by adding the temperature- and speed-dependent engine friction torques.

Figure 3: Physically oriented torque structure: main path with generation of torque and quantity

1 Pressure-control valve, 2 Rail-pressure sensor, 3 Metering unit, 4 High-pressure pump, 5 Crankshaft speed sensor, 6 Boost-pressure sensor, 7 EGR valve, 8 VTG turbocharger, 9 Control flap, 10 Air-mass meter.
n Rotational speed, Q_S Start quantity, t_i Activation period, T Fuel temperature, M Torque, Q Injected fuel quantity, φ_i Crankshaft angle.
AD Activation period, AGR Exhaust-gas recirculation, KST Fuel temperature, LDR Boost pressure, SB Start of injection, RD Rail pressure.



Examples of important functions in the EDC

Control of injection (metering) and of air mass

The required injected fuel quantity is calculated from the necessary internal torque. The required air quantity which must be made available by the air system, especially by supercharging, is calculated from this injected fuel quantity. Among others different physical operating times must be taken into account in the air and fuel systems. The fuel quantity can be adapted at the next firing cylinder while the air quantity can be determined by the operating times in the exhaust-gas turbocharger compressor.

Exhaust-gas recirculation is provided to improve emissions; the setpoint pressure of rail-pressure control depends on the engine operating point as well as on the desired quantity. In actual fact, very many further cross-connections and control operations must be taken into consideration, for example the desired quantity must be limited if the air-system is unable to supply the required air quantity quickly enough (for smoke limitation).

With regard to engine and exhaust-emission behavior the accuracy of injected fuel quantity and injection intervals is subject to closely toleranced requirements. Learning functions are increasingly used here.

Speed governing

Idle-speed control

The purpose of idle-speed control is to adjust a defined set speed at idle when the accelerator pedal is not pressed. This set speed can vary depending on the engine's operating condition; thus, for example, a higher idle speed is set for a cold engine than for a hot engine. The idle speed can also be increased for example when the vehicle system voltage is too low, the air conditioner is switched on or the vehicle is rolling. Because the engine is operated at idle relatively frequently in heavy traffic (e.g., “stop and go” or stopped at traffic lights), the idle speed should for emission and consumption reasons be as low as possible. However, this creates drawbacks for the engine's smooth running and starting performance. Idle-speed control must manage with very strongly fluctuating requirements in relation to the adjustment of the specified set speed.

The power demand of the auxiliary systems driven by the engine is variable within wide limits. The alternator for example draws much more power when the vehicle system voltage is low than when it is high; on top of this are demands on the air-conditioner compressor, the power-steering pump, high-pressure generation for diesel injection, etc. Added to these external load torques is the internal engine friction torque, which is very much dependent on the engine temperature. In the usual torque-controlled systems the torque demands of the auxiliary systems and the engine friction torque can be modeled and used as pilot control. The idle controller (minimum-speed governor) only then has to compensate the model inaccuracies.

Maximum-speed governing

The purpose of maximum-speed governing (also called speed regulation) is to protect the engine against unacceptably high rotational speeds. To this end the engine manufacturer lays down a maximum permissible speed which must not be exceeded for an extended period of time, as this would damage the engine. Speed regulation continuously reduces the command torque above the engine's rated-power operating point. Injection ceases shortly above the maximum engine speed. However, speed regulation must be effected as softly as possible (with a ramp function) so as to prevent jerky speed regulation of the engine when accelerating. This is all the more difficult to achieve the closer together the rated-power operating point and the maximum speed are.

Intermediate-speed regulation

Intermediate-speed regulation is used for commercial vehicles with power take-offs (e.g., crane operation) or for special vehicles (e.g., fire pump). When it is activated, the engine is regulated to an intermediate speed.

The speed can be input here via the cruise-control operating panel or via a communication interface.

Cruise control and road-speed limitation
The vehicle-speed controller enables the vehicle to drive at a constant speed. It regulates the vehicle's speed to a desired value without the driver having to press the accelerator pedal. This value can be adjusted via an operating lever or via steering-wheel buttons. The command torque is increased or decreased until the measured actual speed equals the set desired speed.

Road-speed limitation limits the maximum speed to a settable or legally prescribed value, even when the accelerator pedal continues to be pressed. This is an aid to the driver, who consequently cannot unintentionally exceed speed limits. For this purpose road-speed limitation limits the driver-command torque according to the maximum desired speed.

Surge damping control
In the event of sudden load changes the alteration of the engine torque causes the vehicle drivetrain to undergo bucking oscillations. Vehicle occupants perceive these bucking oscillations to be unpleasant periodic changes in acceleration. Surge damping control reduces these changes in acceleration. This is achieved by means of two separate measures:
– filtering of the torque desired by the driver,
– damping by means of active control.

Smooth-running control, fuel-balancing control
Not all the cylinders of an engine generate the same torque for an identical injection time. This can be down to differences in cylinder compression, differences in cylinder friction or differences in the hydraulic injection components. The consequences of these torque differences are irregular engine running and increased engine emissions. The purpose of smooth-running control or fuel-balancing control is to detect such differences with reference to the resulting speed fluctuations and balance these difference by means of a targeted adaptation of the injected fuel quantity of the cylinder in question.

Altitude correction
Atmospheric pressure decreases at increasing altitude. Consequently the cylinder charge of combustion air becomes smaller. The injected fuel quantity must therefore be reduced. If the same quantity were to be injected as at high atmospheric pressure, there would be heavy smoke emissions due to the air deficiency. The atmospheric pressure is detected by the ambient-pressure sensor in the ECU. In this way the injected fuel quantity can be reduced at high altitudes. The atmospheric pressure also influences boost-pressure control and torque limitation.

Shutdown
The upshot of the operating principle of auto-ignition is that a diesel engine can only be brought to a stop by interrupting the fuel supply. With Electronic Diesel Control the engine is shut down via the ECU input “injected fuel quantity zero” (e.g., no activation of the solenoid valves, closing of the metering unit). A soft shutdown (avoidance of “thumping” during shutdown) can be achieved through the closing of possible throttle valves or exhaust-gas flaps.

λ control
The wideband λ sensor in the exhaust pipe measures the residual oxygen content in the exhaust gas. The air/fuel ratio (excess-air factor λ) can be inferred from this. In a diesel engine (which is always operated with excess air, i.e., $\lambda > 1$), the signal from the λ sensor serves
– to improve the accuracy of exhaust-gas recirculation,
– to deliver full-load smoke limitation,
– to detect unwanted combustion.

Boost-pressure control
Boost-pressure control of the exhaust-gas turbocharger improves the torque characteristic in full-load operation and the charge cycle in part-load operation. The setpoint value for the boost pressure depends on the rotational speed, the injected fuel quantity, the coolant temperature, the air temperature, and the ambient-air pressure. It is compared with the actual value of the boost-pressure sensor. In the event of a control deviation the bypass valve or

the guide vanes of the turbocharger with variable turbine geometry are actuated via the ECU.

Fan activation
The engine control unit activates the engine's fan wheel above a certain engine temperature. The fan wheel continues to be operated for a certain amount of time even after the engine has come to a stop. This after-run time depends on the current coolant temperature and the load condition of the last driving cycle.

Exhaust-gas recirculation
Exhaust gas is directed into the engine's intake tract to reduce the NO_x emissions. This is done via a duct whose cross-section can be varied by an exhaust-gas recirculation valve. The EGR valve is activated for example by an electrical controller. It is difficult to measure the recirculated exhaust-gas flow due to the high temperature and pollutant content of the exhaust gas. Control is therefore effected indirectly via an air-mass meter in the fresh-air mass flow. Its measured value is compared in the ECU with the engine's theoretical air requirement. This is determined from various characteristic data (e.g., engine speed). The lower the actual measured fresh-air mass compared with the theoretical air requirement, the higher the recirculated exhaust-gas content.

Substitute functions
If individual input signals fail, the ECU lacks important information for the calculations. In this case activation is effected with the aid of substitute functions.

Example 1
The fuel temperature is required to calculate the injected fuel quantity. If the fuel-temperature sensor fails, the ECU computes with a substitute value which can be calculated for example using a model from the engine temperature, the ambient temperature, and the tank contents. This must be chosen so as not to give rise to excessive soot formation. This can cause the power to drop in some operating ranges in the event of a faulty fuel-temperature sensor.

Example 2
If the camshaft position sensor fails, the ECU uses the signal from the crankshaft speed sensor as a substitute value. Different vehicle manufacturers have different concepts which are used to determine via the crankshaft-signal characteristic when cylinder 1 is in the compression cycle. The restart, however, takes a little longer as a consequence of these substitute functions.

The different substitute functions may vary from vehicle manufacturer to vehicle manufacturer. All faults are stored by means of the diagnosis function and can be read out in the workshop.

Data exchange with other systems
The engine-management system can communicate via bus systems, such as for example the CAN bus (Controller Area Network), with the ECUs of other vehicle systems and use the information as input signals for control algorithms. The driving speed for example is only calculated in the ESC control unit (Electronic Stability Control) and communicated via the bus to other systems.

Application

Function and meaning of application
Application often also refers to the integration of a component or a subsystem in a higher-level system. Here the generic term application is usually also used for the adaptation of hardware components, while the term calibration is only used for the input of data for the functionalities stored in the ECU.

The application/calibration process of Electronic Diesel Control comprises the matching of different subsystems (e.g., injection system and exhaust-gas treatment system) to the complete system (here: drivetrain or powertrain) with the aim of simultaneously representing the different requirements applicable to the vehicle with regard to driving safety, emissions minimization (in the test cycle and under real driving conditions), durability, engine and component protection, the OBD system (on-board diagnostics) and the engine and vehicle characteristics established by the manufacturer (e.g., torque, max. power, noise, smooth running, response, and performance). In order to optimize these in part opposing objectives, a wealth of characteristic values, program maps, and characteristic curves are calibrated or applied.

The end result is a data record which after extensive testing for series production is stored with write protection in a read-only memory (e.g., flash EPROM). Several 10,000 calibration data are processed in modern engine-management systems. Processing takes place with dedicated tools in the laboratory, on engine and vehicle test benches, on test tracks under real driving conditions and during winter, summer, and altitude trials also under extreme conditions.

References
[1] K. Reif (Editor): Dieselmotor-Management. 6th Edition, Verlag Springer Vieweg, 2019.

Low-pressure fuel supply

The function of the fuel-supply system is to store and filter the required fuel and to make fuel available to the fuel-injection installation at a specific supply pressure under all operating conditions. The fuel return flow is cooled in some applications. The fuel supply system comprises the following essential components (Figures 1 to 3): the fuel tank, the preliminary filter (not for unit-injector system for passenger cars), the presupply pump (optional, also in-tank pump for passenger cars), the fuel filter, the fuel pump (low pressure), the pressure-control valve (overflow valve), the fuel cooler (optional), and low-pressure fuel lines.

Individual components can be combined to form assemblies (e.g. fuel-supply pump with pressure limiter). In axial- and radial-piston distributor pumps and partially in the common-rail system, the fuel-supply pump is integrated into the high-pressure pump.

Essentially, the fuel-supply system differs greatly, depending on the fuel-injection system used, as the following figures for common rail (Figure 1), unit injector (Figure 2) and radial-piston pump (Figure 3) show.

Fuel supply and delivery

The function of the fuel pump in the low-pressure stage is to supply the high-pressure components with sufficient fuel in every operating state, with a low level of noise, at the required pressure, and over the entire service life of the vehicle. Different types are used, depending on the field of application.

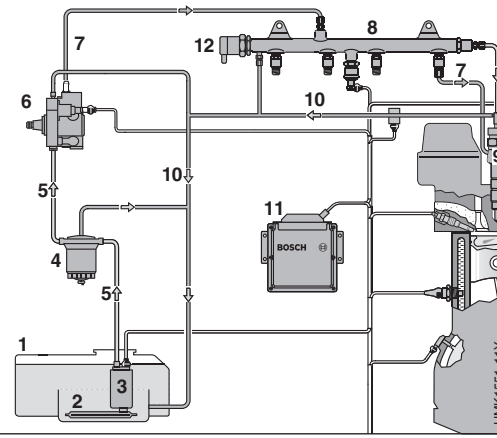
Electric fuel pump

The structure of the electric fuel pump (Figure 4) is comparable to the design variations of the pumps used on spark-ignition engines (see Electric fuel pump). However, there are some different features:

– Roller-cell pumps are usually used for diesel applications (see Roller-cell pump),
– The electric motor of the pump is designed with carbon commutators instead of copper commutators.

Figure 1: Fuel system of a fuel-injection installation with common rail for passenger cars

1 Fuel tank,
2 Preliminary filter,
3 Fuel pump,
4 Fuel filter,
5 Low-pressure fuel lines,
6 High-pressure pump,
7 High-pressure fuel lines,
8 Rail,
9 Injector,
10 Fuel return line,
11 ECU,
12 Pressure-control valve.



Use of electric fuel pumps:
– Optional for distributor injection pumps (only in event of long fuel lines or great difference in height between fuel tank and fuel-injection pump)
– Unit-injector system (passenger cars)
– Common-rail system (passenger cars)

Gear pump

The gear pump is mounted directly onto the engine or, in the case of a common rail, it is integrated into the high-pressure pump. It is driven mechanically by way of a clutch, a gear wheel, or a toothed belt.

The primary components are two intermeshing, counter-rotating gear wheels (Figure 5) which deliver the fuel into the

Figure 2: Fuel system of a fuel-injection installation with unit-injector system for passenger cars

1 Fuel tank with presupply pump and prefilter,
2 Fuel cooler,
3 Electronic control unit,
4 Fuel filter,
5 Fuel feed line,
6 Fuel return line,
7 Tandem pump as fuel pump,
8 Fuel-temperature sensor,
9 Sheathed-element glow plug,
10 Injector.

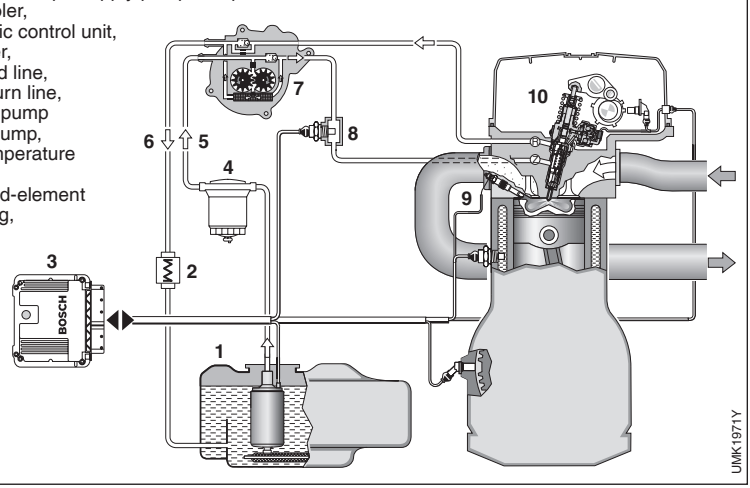
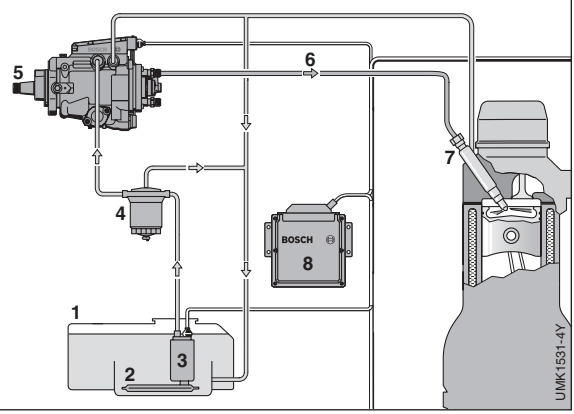


Figure 3: Fuel system of a fuel-injection installation with radial-piston fuel-injection pump for passenger cars

1 Fuel tank,
2 Preliminary filter,
3 Presupply pump,
4 Fuel filter,
5 Radial-piston distributor injection pump with integrated fuel pump,
6 High-pressure line,
7 Nozzle-holder assembly,
8 Control unit.



tooth spaces from the suction side to the pressure side. The line of contact of the gear wheels provides the seal between the suction and pressure sides and prevents fuel from flowing back.

The delivery quantity is roughly proportional to the engine speed. At high rotational speeds, therefore, the quantity is limited by throttle action on the intake side or by an overflow valve on the pressure side, which discharges the feed volume to the inlet inside the pump. This prevents unnecessarily high flows in the vehicle's conduit system.

Applications:
– For single-cylinder pump systems for commercial vehicles (unit-injector and unit-pump system, individual injection pump (PF))
– for common-rail systems (commercial vehicles, in part for passenger cars).

Figure 4: Single-stage electric fuel pump
A Pump element,
B Electric motor,
C Connecting cover.
1 Pressure side,
2 Motor armature,
3 Pump element,
4 Pressure limiter,
5 Suction side,
6 Non-return valve.

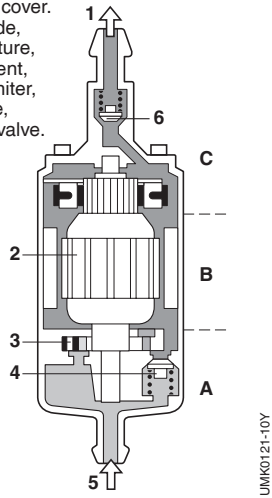
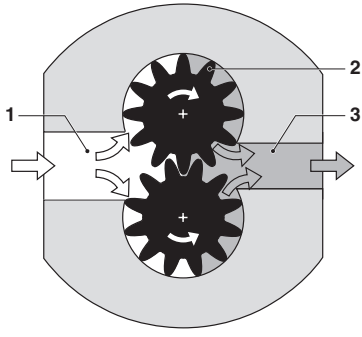


Figure 5: Gear fuel pump (diagram)
1 Suction side,
2 Drive gear,
3 Pressure side.

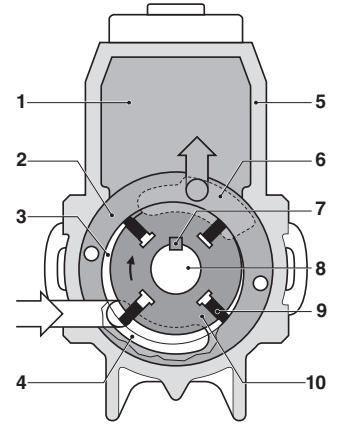


Vane-type supply pump

The vane-type supply pump (Figure 6) is mounted on the drive shaft in the distributor injection pump. The impeller is centered on the drive shaft and supported by a Woodruff key. An eccentric ring mounted in bearings in the housing surrounds the impeller.

The fuel passes through the inlet passage and a kidney-shaped recess into the space created by the impeller, the

Figure 6: Vane pump
1 Inner pump chamber,
2 Eccentric ring,
3 Sickle-shaped cell,
4 Fuel inlet (suction cell),
5 Pump housing,
6 Fuel outlet (pressure cell),
7 Woodruff key,
8 Drive shaft,
9 Vane, 10 Impeller.



vane and the eccentric ring. The centrifugal force generated by the motion of rotation presses the four vanes of the impeller outwards against the eccentric ring. The fuel between the underside of the vanes and the impeller supports this outward movement of the vanes. The rotary motion forces the fuel located between the vanes into the upper kidney-shaped recess and through a bore to the outlet.

Application: Integrated presupply pump in the distributor injection pump.

Locking-vane pump

In a locking-vane pump (Figure 7), springs press two locking vanes against a rotor. As the rotor turns, the volume is increased on the suction side and fuel is drawn into two chambers. The volume is reduced on the pressure side and fuel is delivered from the two chambers. The locking-vane pump delivers fuels at very low speeds.

Application: Unit-injector system for passenger cars.

Tandem fuel pump

The tandem fuel pump (Figure 8) is an assembly consisting of a fuel pump and a vacuum pump, for example, for the brake booster. From a functional perspective, these are two separate pumps, which, however, are driven by a common drive shaft. It is integrated into the engine cylinder head and driven by the engine camshaft.

The fuel pump itself is either a locking-vane pump or a gear pump, and even at low speeds (cranking speeds) delivers enough fuel to ensure that the engine starts reliably. The pump's delivery quantity is proportional to the rotational speed.

The pump contains a variety of valves and throttling orifices. The pump's maximum delivery quantity is limited by the suction throttling orifice so that not too much fuel is delivered. The pressure-relief valve limits the maximum pressure in the high-pressure stage. Vapor bubbles in the fuel feed are eliminated in the fuel-return throttling bore. If there is air in the fuel system (e.g. if the vehicle has been driven until the fuel tank is empty), the low-pressure control valve remains

Figure 7: Locking-vane pump (diagram)
1 Rotor,
2 Suction side (inlet),
3 Spring,
4 Locking vane,
5 Pressure side.

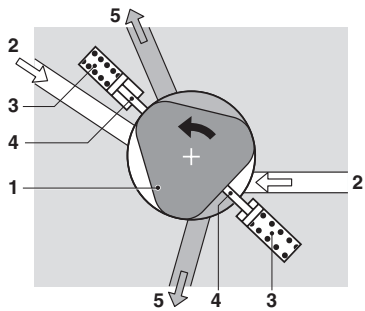


Figure 8: Fuel pump in a tandem pump
1 Return to tank,
2 Supply from tank,
3 Pump element (gear wheel),
4 Throttling bore,
5 Filter,
6 Suction throttle,
7 Pressure-relief valve,
8 Connection for pressure measurement,
9 Supply, injector,
10 Return, injector,
11 Non-return valve,
12 Bypass.

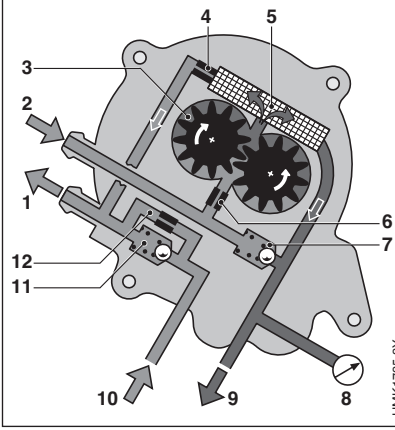


Figure 9: Pressure-control valve
1 Valve body,
2 Screw,
3 Compression spring,
4 Edge seal,
5 Accumulator plunger,
6 Accumulator volume,
7 Conical seat.

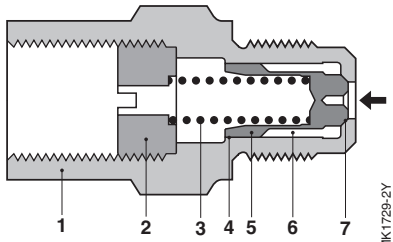
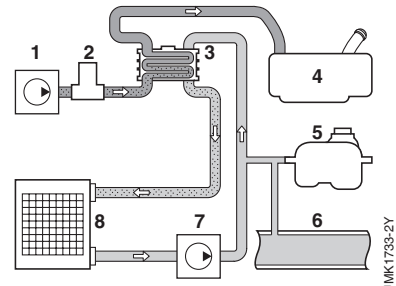


Figure 10: Fuel-cooling circuit
1 Fuel-supply pump,
2 Fuel-temperature sensor,
3 Fuel cooler,
4 Fuel tank,
5 Expansion tank,
6 Engine-cooling circuit,
7 Coolant pump,
8 Auxiliary cooler.



closed. The air is forced out of the fuel system through the bypass by the pressure of the pumped fuel.

Application: Unit-injector system for passenger cars.

Low-pressure control valve
The pressure-control valve (Figure 9, also called overflow restriction) is built into the fuel return, for common-rail usually into the high-pressure pump. It provides an adequate operating pressure in the low-pressure stage of the injection systems under all operating conditions and ensures that the pumps are filled uniformly.

For example, the accumulator plunger of the valve for the unit injector and unit pump opens at the initial opening pressure of 300 to 350 kPa (3 to 3.5 bar). A compression spring compensates for minor pressure variations in accumulator volume. At an opening pressure of 4 to 4.5 bar, an edge seal opens and produces a noticeable increase in flow rate.

There are two screws with different spring-force settings for presetting the opening pressure.

Fuel cooler
The high pressure in the injector of the unit-injector system for passenger cars and some common-rail systems heats up the fuel to such an extent that it has to be cooled before it flows back. The fuel returning from the injector flows through the fuel cooler (heat exchanger) and dissipates thermal energy to the coolant in the fuel-cooling circuit. This is separate from the engine-cooling circuit because the coolant temperature of an engine running at operating temperature is too high to cool the fuel. The fuel-cooling circuit is connected to the engine-cooling circuit at the compensating reservoir. This helps to keep the fuel-cooling circuit filled and compensates for any changes in volume caused by temperature fluctuations (Figure 10).

In part, common-rail uses other cooling concepts (such as fuel-air heat exchangers on the floor of the vehicle). Due to the reduced power loss, however, newer common-rail systems with quantity-controlled high-pressure pumps are largely being used without fuel coolers.

Diesel filter

Function

As in spark-ignition (SI) engines, it is essential in a diesel engine to ensure that the fuel system is protected against contamination. Contaminants are introduced into the system during refueling or can get into the tank via the tank ventilation and thus into the fuel itself. The function of the fuel filter is to reduce contamination from particulates so as to protect the fuel-injection system.

Compared with gasoline-injection systems, diesel fuel-injection systems require increased wear protection and finer filters to cater for the much higher injection pressures. Diesel fuel is also more heavily contaminated than gasoline.

Design

Diesel filters are designed as easy-change filters (Figure 1). Spin-on filters, in-line filters and nonmetal filter elements are widely used as replacement parts in aluminum, plastic or sheet-steel filter cases (to comply with increased requirements in crashes). Spiral vee-shape filter elements are the preferred choice. The diesel fuel filter is installed in the low-pressure circuit: in suction systems upstream of the fuel-supply pump and in pressure systems downstream of the electric fuel pump. The trend is now towards pressure systems.

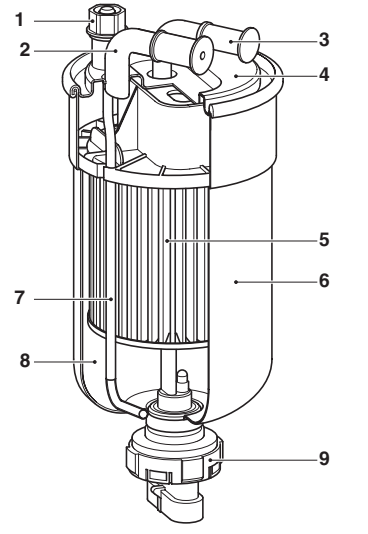
Requirements

In the last few years, the requirements for filter fineness have become more stringent owing to the introduction of common-rail systems with higher injection pressures and advanced unit-injector systems for passenger cars and commercial vehicles. Depending on the application (fuel contamination, periods of engine standstill), the new systems require a filtration efficiency of between 85 % and 98.6 % (particulate interval 3 to 5 µm, ISO/TS 13353 [1] and ISO 19438 [2]). The fuel filters fitted to the latest generation of automobiles must be capable of storing larger volumes of particulates due to the longer service intervals, and of separating superfine particulates efficiently. This can only be achieved by using special filter media, e.g. consisting of multiple layers made of synthetic microfiber. These filter media exploit the effects of a fine prefilter, and guarantee maximum particulate-retention capability by separating particulates inside each filter layer.

Typical replacement intervals today are between 60,000 and 90,000 km. These intervals are significantly shortened in markets where the diesel quality is poor, such as Eastern Europe, China,

Figure 1: Diesel line filter

1 Water drain plug,
2 Fuel inlet,
3 Fuel outlet,
4 Filter cover with galvanized coating,
5 Radial vee-form filter element with two-ply filter medium,
6 Compression-proof housing made from Galfan (hot-dip-modified sheet with a zinc-aluminum alloy coating on both sides),
7 Tube to drain,
8 Water accumulation chamber,
9 Water-level and temperature sensors.



India, and the USA. Likewise, when biodiesel is used, halving the replacement interval is recommended.

Water separation

A second essential function of the diesel filter is to separate emulsified and free water in order to prevent corrosion damage. An effective water separation of more than 93 % at rated flow (testing as per ISO 4020 [3]) is particularly important for distributor injection pumps and common-rail systems. Water separation takes place by coalescence on the filter medium (formation of droplets due to the different surface tensions of water and fuel). Separated water collects in the chamber at the bottom of the filter housing (Figure 1). Conductivity sensors are used in some cases to monitor the water level. The water is drained manually by opening a water drain plug or pressing a pushbutton switch.

For extra heavy-duty requirements, it is beneficial to fit an additional preliminary filter with water separator on the suction or pressure side. Preliminary filters of this type are mostly used for commercial vehicles in countries where diesel quality is poor.

Modular additional functions

New-generation diesel fuel filters integrate modular additional functions such as, for example, fuel preheating to prevent paraffin clogging during winter operation. Fuel preheating can be effected electrically or by utilizing the hot fuel return flow from the engine. In the first instance, PTC heaters (Positive Temperature Coefficient) are installed in the filter. The second instance requires the fitting of a bimetal valve or a wax expansion element which opens at low temperatures and allows fuel to flow back to the filter.

Further additional functions are maintenance indication by way of differential-pressure measurement and filling and venting devices.

References

[1] ISO/TS 13353:2002: Diesel fuel and petrol filters for internal combustion engines – Initial efficiency by particle counting.

[2] ISO 19438:2003: Diesel fuel and petrol filters for internal combustion engines – Filtration efficiency using particle counting and contaminant retention capacity.

[3] ISO 4020:2001: Road vehicles – Fuel filters for diesel engines – Test methods.

Common-rail fuel-injection system

System overview

Requirements

The demands placed on diesel-engine fuel-injection systems are continually increasing. Higher injection pressures, faster injector switching times, and a variable rate-of-discharge curve modified to the engine operating state have made the diesel engine economical, clean, and powerful.

The main advantage of the common-rail system (CR) over conventional fuel-injection systems is its ability to vary injection pressure and timing over a broad scale. This was achieved by separating pressure generation (in the high-pressure pump) from the fuel-injection system (injectors). The rail here acts as a pressure accumulator (Figure 1).

The common-rail system is a highly flexible system for adapting fuel injection to the engine. This is achieved by an injection pressure adapted to the operating status of 200 to 2,500 bar, a variable start of injection, and the possibility of using pre-injection and secondary-injection events (even highly retarded secondary-injection events are possible). In this way, the common-rail system helps to raise specific power output, lower fuel consumption, reduce noise emission, and decrease pollutant emission in diesel engines.

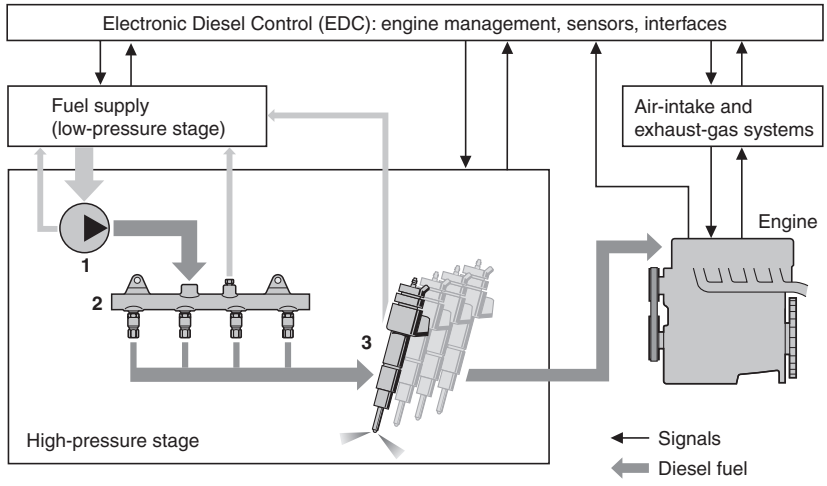
Common rail is today the most commonly used fuel-injection system for modern passenger-car and commercial-vehicle diesel engines.

Design

The common-rail system consists of the following main component groups (Figure 1): the low-pressure system comprising the components of the fuel-supply system, the high-pressure system comprising the high-pressure pump, the rail, the injectors and the high-pressure fuel lines, and Electronic Diesel Control (EDC).

The key components of the common-rail system are the injectors connected to the rail. They are fitted with a rapid-action valve (solenoid valve or piezo actuator) which opens and closes the nozzle. This

Figure 1: System areas of engine management with the common-rail fuel-injection system
1 High-pressure pump, 2 Rail, 3 Injectors.



permits control of the injection process for each cylinder.

Fuel supply
In common-rail systems for passenger cars and light-duty commercial vehicles, electric fuel pumps or gear pumps are used to deliver fuel to the high-pressure pump. It is exclusively gear pumps that are used for heavy-duty commercial vehicles (see Low-pressure fuel supply).

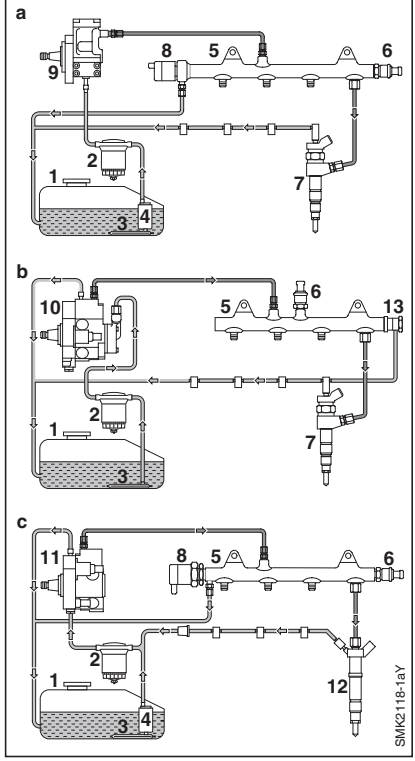
Systems with electric fuel pump
Systems with an electric fuel pump are increasingly gaining acceptance in the passenger-car market. There are several reasons for this:
– Using a controlled electric fuel pump results in a lower power demand (CO_2 advantage).
– As well as supplying the fuel-injection system, an electric fuel pump permits in parallel the operation of suction jet pumps in the tank; these pumps fill the swirl pot and thus ensure the presence of fuel in the intake area (see Fuel-delivery module).
– The fact that the pressure build-up in the low-pressure system is not dependent on engine speed improves the starting response.

The electric fuel pump is usually part of the in-tank unit (in the fuel tank) or occasionally is fitted in the fuel line (in-line). It intakes fuel via a pre-filter and delivers it to the high-pressure pump at a pressure of 4 to 6 bar (Figures 2a and 2c). The maximum delivery rate is, depending on engine performance, 150 to 240 *l*/h.

Systems with gear pump
The gear pump is flanged to the high-pressure pump and is driven by its input shaft (Figure 2b). In this way, the gear pump starts delivery only after the engine has started. Delivery rate is dependent on the engine speed and reaches rates up to 400 *l*/h at pressures up to 7 bar.

Combination systems
There are also applications where the two pump types are used. The electric fuel pump improves starting response, in

Figure 2: Common-rail systems for passenger cars (examples)
a) Continuous-delivery radial-piston pump: pressure control on the high-pressure side with pressure-control valve.
b) Demand-controlled radial-piston pump: pressure control on the suction side with metering unit flanged to the high-pressure pump.
c) Demand-controlled radial-piston pump: two-actuator system with pressure control on the suction side via the metering unit and pressure control on the high-pressure side via the pressure-control valve.
1 Fuel tank, 2 Fuel filter,
3 Preliminary filter, 4 Electric fuel pump,
5 Rail, 6 Rail-pressure sensor,
7 Solenoid-valve injector,
8 Pressure-control valve,
9 Continuous-delivery high-pressure pump,
10 Demand-controlled high-pressure pump with mounted gear presupply pump and metering unit,
11 Demand-controlled high-pressure pump with metering unit,
12 Piezo injector,
13 Pressure limiter.

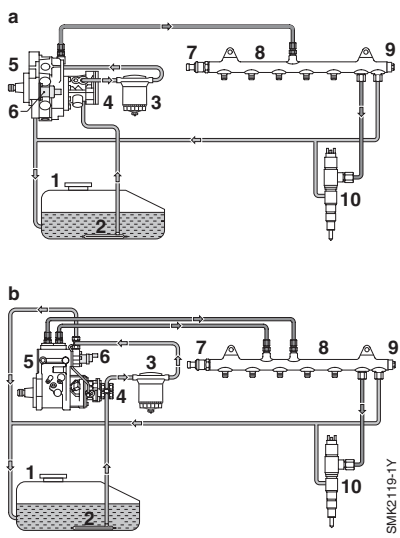


particular for hot starts, since the delivery rate of the gear pump is lower when the fuel is hot, and therefore, thinner, and at low pump speeds. In addition, the electric fuel pump ensures the operation of suction jet pumps in the tank.

Combination systems are however rarely used, due to the increased costs and the increased power demand for two pumps.

Figure 3: Common-rail systems for commercial vehicles (examples)

a) Demand-controlled radial-piston pump with pressure control on the suction side via the metering unit,
b) Demand-controlled two-plunger in-line pump with pressure control on the suction side via the metering unit.
1 Fuel tank,
2 Preliminary filter,
3 Fuel filter,
4 Gear presupply pump,
5 High-pressure pump,
6 Metering unit,
7 Rail-pressure sensor,
8 Rail,
9 Pressure limiter,
10 Injector.



Fuel filtering

In passenger-car systems with an electric fuel pump the fuel filter is located on the pressure side between the electric fuel pump and the high-pressure pump. In systems with a gear pump the fuel filter is fitted between the fuel tank and the gear pump. In contrast, the fuel filter (fine filter) is fitted on the pressure side in commercial-vehicle systems. For this reason, an exterior fuel inlet is required, in particular when the gear pump is flanged to the high-pressure pump (Figure 3).

Pressure generation

In the common-rail pressure-accumulator fuel-injection system, the functions of pressure generation and fuel injection are separate. The injection pressure is generated independent of the engine speed and the injected fuel quantity. Electronic Diesel Control controls each of the components.

The functions of pressure generation and fuel injection are separated by the rail's accumulator volume. The pressurized fuel is held in the accumulator volume ready for injection.

A continuously operating high-pressure pump driven by the engine produces the desired injection pressure. Pressure in the fuel rail is maintained irrespective of engine speed or injected fuel quantity.

The high-pressure pump is a radial-piston pump. On commercial vehicles, in contrast to passenger cars, an in-line pump is sometimes fitted.

Pressure control

Different procedures are used for pressure control.

Control on the high-pressure side

The desired rail pressure is regulated on the high-pressure side by a pressure-control valve (Figure 2a). Fuel not required for injection flows back to the low-pressure circuit via the pressure-control valve. This type of control loop allows rail pressure to react rapidly to changes in operating point (e.g. in the event of load changes).

Fuel-delivery control on the suction side
Another way of controlling rail pressure is to control fuel delivery on the suction side (Figures 2b and 3). The metering unit flanged on the high-pressure pump makes sure that the pump delivers exactly the right quantity of fuel to the fuel rail in order to maintain the injection pressure required by the system. In a fault situation, the pressure-relief valve prevents rail pressure from exceeding a maximum. Fuel-delivery control on the suction side reduces the quantity of fuel under high pressure and lowers the power input of the pump. This has a positive impact on fuel consumption. At the same time, the temperature of the fuel flowing back to the fuel tank is reduced in contrast to the control method on the high-pressure side.

Two-actuator system
The two-actuator system (Figure 2c) combines pressure control on the suction side via the metering unit and control on the high-pressure side via the pressure-control valve, thus marrying the advantages of suction-side fuel-delivery control with the dynamic response of control on the high-pressure side. Another advantage compared with control on the low-pressure side only is that the high-pressure side is also controllable when the engine is cold. The high-pressure pump then delivers more fuel than is injected and pressure is controlled by the pressure-control valve. Compression heats the fuel, thus eliminating the need for an additional fuel heater.

Fuel injection
The injectors spray fuel directly into the engine's combustion chambers. They are supplied by short high-pressure fuel lines connected to the fuel rail. The engine control unit controls the switching valve integrated into the injector to open and close the injector nozzle.

The injector opening times and system pressure determine the quantity of fuel delivered. At a constant pressure, the fuel quantity delivered is proportional to the switching time of the solenoid valve. This is, therefore, independent of engine and pump speeds (time-based fuel injection).

Hydraulic performance potential
Separating the functions of pressure generation and fuel injection opens up further degrees of freedom in the combustion process compared with conventional fuel-injection systems; the injection pressure is more or less freely selectable within the program map.

The common-rail system allows a further reduction in exhaust-gas emissions by introducing pre-injection events and multiple injection events and also attenuating combustion noise significantly. Multiple injection events of up to seven per injection cycle can be generated by triggering the rapid-action switching valve several times. Hydraulic pressure is used to augment nozzle-needle closing, ensuring rapid termination of the injection process.

Control and regulation
Operating principle
The engine control unit detects the accelerator-pedal position and the current operating states of the engine and vehicle by means of sensors (see "Electronic Diesel Control"). The data collected includes: the crankshaft speed and the crankshaft angle, the rail pressure, the charge-air pressure, the intake-air, coolant and fuel temperature, the inducted air mass, and the wheel speed (for calculating the driving speed). The electronic control unit evaluates the input signals. In sync with combustion, it calculates the triggering signals for the pressure-control valve or the metering unit, the injectors, and the other actuators (e.g. the EGR valve, exhaust-gas turbocharger actuators, etc.).

Basic functions
The basic functions involve the precise control of diesel-fuel injection timing and fuel quantity at the reference pressure. In this way, they ensure that the diesel engine has low consumption and smooth running characteristics.

Correction functions for calculating injection time
A number of correction functions are available to compensate for tolerances between the fuel-injection system and the engine.

Fuel-balancing control
These tolerances result in different torque build-up of individual cylinders. The consequences of these torque differences are irregular engine running and increased exhaust-gas emissions. Injected-fuel-quantity corrections are calculated from the resulting speed fluctuations. Smooth running is increased by specifically adapting the injection time for each cylinder (smooth-running control).

Injector delivery compensation
Injector delivery compensation enables the fuel-quantity deviations of brand-new injectors to be corrected. To this end a wide range of measured data is recorded during injector production for each injector and then applied to the injector in the form of a data matrix code. Information about the lift behavior is also added for a piezo in-line injector. These test data are transmitted during vehicle production to the engine ECU. These values are used during engine operation to compensate deviations in metering and switching performance.

Zero delivery calibration
To achieve both noise reduction and emission targets it is particularly important to be safely in control of small pre-injections over the service life of the vehicle. Injector fuel-quantity drifts must therefore be compensated. To this end a small quantity of fuel is specifically injected into a cylinder in overrun conditions. The speed sensor detects the resulting torque increase as a small dynamic change in speed. This torque increase which is not felt by the driver is clearly linked to the injected fuel quantity. The process is repeated in succession for all the cylinders and for different operating points. A learning algorithm establishes the smallest changes in the pre-injection quantity and corrects the activation duration for the injectors accordingly for all the pre-injections.

Fuel-quantity mean-value adaptation
The deviation of the actually injected fuel quantity from the setpoint value is required to correctly adapt exhaust-gas recirculation and charge-air pressure. Fuel-quantity mean-value adaptation determines the value for the fuel quantity averaged over all the cylinders from the signals from the λ sensor and the air-mass meter. Correction values for the injected fuel quantity are calculated from the comparison of setpoint and actual values.

Supplementary functions
Additional open- and closed-loop control functions perform the tasks of reducing exhaust-gas emissions and fuel consumption, or providing added safety and convenience. Examples of such functions are control of exhaust-gas recirculation, boost-pressure control, cruise control, and the electronic vehicle immobilizer.

The integration of Electronic Diesel Control in a vehicle's overall system network facilitates the exchange of data with, for example, transmission control or air conditioning. A diagnosis interface facilitates the analysis of stored system data when the vehicle is serviced.

Injectors

Solenoid-valve injectors

Design and operating principle

The solenoid-valve injector can be subdivided into a number of function modules: the injection nozzle, the hydraulic servo valve to actuate the valve plunger, the nozzle needle, and the solenoid valve.

The fuel is routed from the high-pressure port (Figure 4a) via an inlet passage to the injection nozzle and via the inlet restrictor into the valve control chamber. The valve control chamber is connected by the outlet restrictor, which can be opened by the solenoid valve, to the fuel return.

The function of the injector can be subdivided into four operating states when the engine and the high-pressure pump are operating. The operating states are caused by the balance of forces acting on the injector components. The nozzle spring closes the nozzle when the engine is not running and there is no pressure in the rail.

All the solenoid-valve injectors have a hydraulic port on the solenoid group for the return flow to the low-pressure system. The return flow is made up of the control quantity (only while the valve is open) and the leakage of the guides at the nozzle, the valve plunger, and the solenoid valve when a pressure-compensated valve is used (see Injector variants). To increase hydraulic efficiency, injectors with a maximum injection pressure greater than 2,000 bar have no leakage at the nozzle guide and valve plunger.

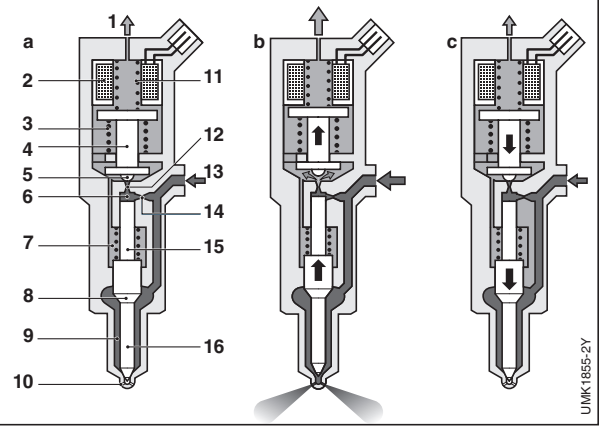
Nozzle closed (rest position)

In its rest position, the solenoid valve is not triggered (Figure 4a). The solenoid valve, in response to the force of the solenoid-valve spring, closes the valve seat, thereby stopping the flow through the outlet restrictor. Inside of the valve control chamber, the pressure rises to the pressure in the fuel rail. The same pressure is also present in the nozzle chamber. The forces acting on the end faces of the control plunger resulting from the fuel-rail pressure and the force of the nozzle spring hold the nozzle needle in the closed position against the opening force acting on the pressure shoulder of the needle.

Figure 4: Functioning principle of solenoid-valve injector (schematic representation)

a) Nozzle closed (rest position), b) Nozzle opens (start of injection),
c) Nozzle closes (end of injection).

1 Fuel return, 2 Solenoid coil (solenoid valve), 3 Overstroke spring, 4 Solenoid armature, 5 Valve seat, here with pressure-loaded ball valve, 6 Valve control chamber, 7 Nozzle spring, 8 Pressure shoulder of nozzle needle, 9 Chamber volume of injection nozzle, 10 Nozzle-body seat with injection orifices, 11 Solenoid-valve spring, 12 Outlet restrictor, 13 High-pressure port, 14 Inlet restrictor, 15 Valve plunger (control plunger), 16 Injection-nozzle needle.



Nozzle opens (start of injection)
During the opening phase of triggering (Figure 7) the magnetic force of the electromagnet exceeds the spring force of the solenoid-valve spring. The armature opens the valve completely in the ensuing pick-up-current phase. This releases the flow through the outlet restrictor (Figure 4b).

The rapid opening of the solenoid valve and the required rapid switching times are achieved by controlling solenoid-valve triggering in the ECU at high voltages and currents (Figure 7). After a short time, the increased pickup current is reduced to a lower holding current.

When the outlet restrictor opens, fuel flows from the valve control chamber to the armature chamber and then via the fuel-return line to the fuel tank. The pressure in the valve control chamber drops. Fuel flows continuously through the inlet restrictor into the control chamber. This prevents the pressure in the control chamber from dropping completely. The flows of the outlet and inlet restrictors are adapted to the dynamic response (switching times) of the solenoid valve. The pressure at the nozzle needle remains at the level of that in the rail. The reduction in pressure in the valve-control chamber reduces the force acting on the control plunger and opens the nozzle needle. Fuel injection commences.

Nozzle open
The rate of movement of the nozzle needle is determined by the difference in the flow rates through the inlet and outlet restrictors and by the area ratios at the nozzle needle and valve plunger. Fuel is injected into the combustion chamber at a pressure approaching that in the fuel rail.

The balance of forces in the injector is similar to that during the opening phase. At a given rail pressure, the fuel quantity injected is proportional to the length of time that the solenoid valve is open. This is entirely independent of the engine or pump speed (time-based injection system).

Nozzle closes (end of injection)
When the solenoid valve is no longer triggered, the solenoid-valve spring presses the armature down, closes the valve seat, and thus stops the flow through the outlet restrictor (Figure 4c). When the outlet restrictor closes, pressure in the control chamber rises again to that in the fuel rail via the inlet restrictor. The higher pressure exerts a greater force on the control plunger. The force on the valve-control chamber and the nozzle-spring force then exceed the force acting on the nozzle needle, and the nozzle needle moves in the direction of the nozzle-body seat. The flow from the inlet restrictor and the area ratios at the nozzle needle and valve plunger determine the speed with which the nozzle needle closes. The fuel-injection cycle comes to an end when the nozzle needle is resting against its seat, thus closing off the injection orifices.

This indirect method is used to trigger the nozzle needle by means of a hydraulic servo system because the forces required to open the nozzle needle rapidly cannot be generated directly by the solenoid valve. The control quantity required in addition to the injected fuel quantity passes through the valve control chamber restrictors into the fuel return.

Multiple injections of up to ten injection pulses are possible (pilot injection, main injection, secondary injection) within one injection cycle with the latest injector types. The minimum possible time interval is approximately 150 µs.

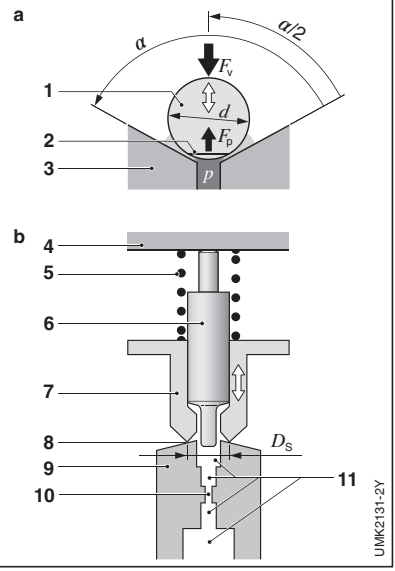
Injector variants
A distinction is made between two different valve concepts with solenoid-valve injectors:
– Injectors with a pressure-loaded ball valve (valve forces act against the applied rail pressure),
– Injectors with a pressure-compensated valve (valve forces are virtually independent of the rail pressure).

Pressure-loaded ball valve
In the case of a pressure-loaded ball valve, the pressure of the compressed fuel acts on the area resulting from valve-seat angle and ball diameter (Figure 5a, see also Figure 4). This pressure generates an opening force. The spring force must be at least of sufficient magnitude that valve remains closed in the non-active state. In practice, the spring force is approximately 15% greater than the hydraulic force at maximum injection pressure in order on the

one hand to achieve a sufficient dynamic response when the valve closes, and on the other hand to ensure sufficient leak-tightness when closed.

Pressure-compensated valve
In the case of a pressure-compensated valve, there is no area on which the pressure acts and a force can be generated in the opening direction (Figure 5b).

Figure 5: Valves of injector variants
a) Pressure-loaded ball valve,
b) Pressure-compensated valve.
1 Valve ball with diameter d,
2 Valve-seat diameter D, $D = d \sin(90° - \alpha/2)$,
3 Valve section with valve sealing seat,
4 Counter-bearing,
5 Solenoid-valve spring,
6 Armature pin in solenoid assembly (diameter D_S),
7 Solenoid armature,
8 Valve seat (diameter D_S),
9 Valve section with outlet restrictor and valve chamber,
10 Outlet restrictor,
11 Valve chamber.
α Valve-seat angle,
p Rail pressure,
F_p Hydraulic force ($F_p = \pi/4 \cdot D^2 p$),
F_V Valve spring force.



Bosch offers solenoid-valve injectors for passenger cars and light commercial vehicles up to 1,800 bar with a pressure-loaded ball valve. To meet the highest demands in modern diesel engines pressure-compensated valves are used from 1,800 bar, since a maximum valve opening cross-section can also be attained at these high pressures. This improves the hydraulic stability of the servo valve with which the needle lift is controlled. In addition, the valve lift is also significantly reduced at high rail pressures in order also to achieve the required dynamic response for injection pressures in excess of 1,600 bar and to reduce the sensitivity of the valve dynamic response to external influences. The required minimum time intervals between two injections are achieved on account of the high dynamic response. The power demand for triggering the solenoid valves can also be reduced.

These injector variants with a pressure-compensated valve have a potential for rail pressures in excess of 2,000 bar. There is also the possibility of damping pressure pulsations in the injector with a mini-volume and thereby increasing metering precision in the case of multiple injections.

Further development of solenoid-valve injectors
The further development of solenoid-valve injectors is currently being driven in two directions:
– higher maximum permissible injection pressures,
– improving metering precision and reducing the injected-fuel-quantity tolerances.

Higher injection pressures
By increasing the injection pressure, it is possible, with the same engine power, to improve combustion by reducing the injection-orifice diameter at the nozzle and to lower the untreated emissions that arise in the combustion chamber. Solenoid-valve injectors are currently available for a maximum pressure of 2,500 bar.

Metering precision
To improve the metering precision of the injected fuel quantity, sensors are available with which the opening time of the valve and the nozzle needle can be measured and a control loop can be established with software programmed in the engine ECU. Then the opening time of the solenoid valve and consequently the opening time of the nozzle needle can be regulated via the activation period. In this way, it is possible to eliminate to a large extent the deviation of the injected fuel quantity through external influences (e.g., fuel temperatures and viscosity) and through influences in the injector (e.g., wear on the components).

In Bosch solenoid-valve injectors this sensor is integrated in the solenoid of the pressure-compensated valve (Figure 6). The pressure in the valve control chamber is transferred via the armature pin and the pressure distributor to a piezo sensor; the piezo sensor uses the pressure force to generate an electrical signal. This is transmitted via an additional pin in the injector connector to the engine ECU, where it is evaluated.

The pressure characteristic in the control chamber correlates with the opening of the valve and the closing of the nozzle needle. In this way, a control loop can be established to regulate the injected fuel quantities.

Figure 6: Solenoid-valve injector with piezo sensor in the solenoid assembly
1 Valve control chamber,
2 Piezo-sensor element,
3 Discharge fitting,
4 Pressure distributor,
5 Electrical connection (three pins),
6 Solenoid sleeve,
7 Solenoid retaining nut,
8 Solenoid core with coil,
9 Mini-rail volume (high pressure),
10 Extrusion coating,
11 Nozzle needle,
12 Valve spring,
13 Armature pin,
14 Diffuser geometry.

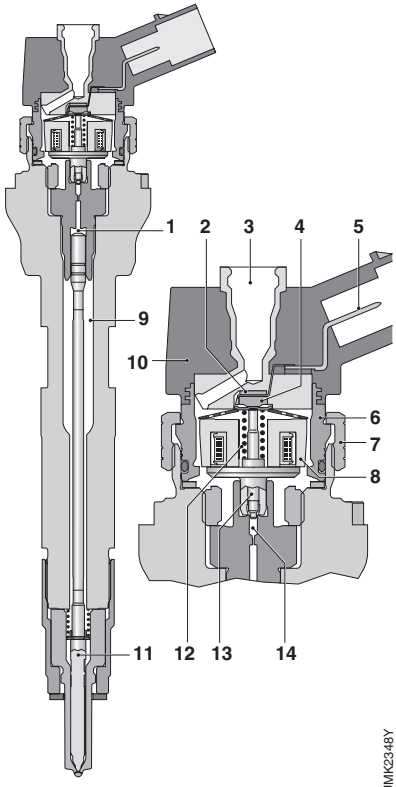
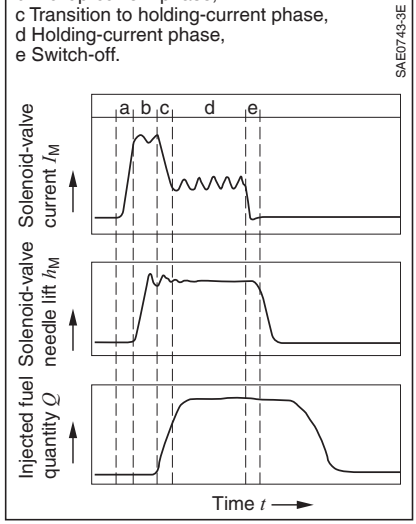


Figure 7: Triggering sequence of a high-pressure solenoid valve for a single injection event
a Opening phase,
b Pickup-current phase,
c Transition to holding-current phase,
d Holding-current phase,
e Switch-off.



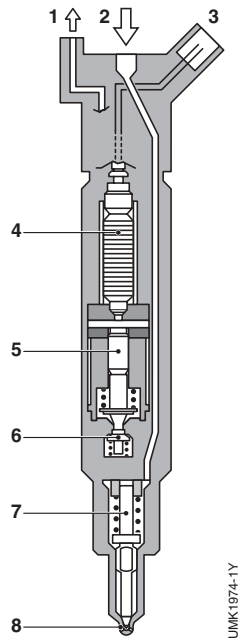
Triggering the solenoid-valve injector
Triggering of the solenoid valve is divided into five phases (Figure 7; the following details apply to injectors with a pressure-loaded ball valve; small values are sufficient for pressure-compensated valves).

Initially, in order to ensure tight tolerances and high levels of reproducibility for the injected fuel quantity, the current for opening the solenoid valve features a steep, precisely defined flank and increases rapidly up to approx. 20 A (opening phase). This is achieved by means of a booster voltage of up to 50 V which is generated in the ECU and stored in a capacitor (boost-voltage store). When this voltage is applied across the solenoid valve, the current increases several times faster than it does when only battery voltage is used. During the pickup-current phase, battery voltage is applied to the solenoid valve. Current control limits pickup current to approx. 20 A.

In order to reduce power loss in the ECU and injector, the current is dropped to approx. 13 A in the holding-current phase. The energy which becomes available when pickup current is reduced to holding current is routed to the booster-voltage capacitor.

The solenoid valve is closed when the current is switched off, and energy is also released in the process. This energy is also routed to the booster-voltage capacitor. The difference between the energy drawn from the booster-voltage capacitor and the energy returned to it is routed to the booster-voltage capacitor via a step-up chopper integrated in the ECU from the vehicle electrical system. Recharging from the vehicle electrical system is performed until the original voltage level required to open the solenoid valve is reached.

Figure 8: Structural design of piezo inline injector
1 Fuel return,
2 High-pressure port,
3 Electrical connection,
4 Piezo actuator,
5 Hydraulic coupler,
6 Servo valve (control valve),
7 Nozzle module with nozzle needle,
8 Injection orifice.



Piezo inline injector
Design and requirements
The design of the piezo inline injector is divided into its main modules in the schematic (Figure 8):
– actuator module (piezo actuator and encapsulation, contact, components for support and force transmission of the actuator),
– hydraulic coupler,
– servo valve (control valve), and
– nozzle module.

A direct response of the nozzle needle to piezo-actuator operation is achieved by coupling the servo valve closely to the nozzle needle. The delay between the electric start of triggering and hydraulic response of the nozzle needle is about 150 µs. At the same time, on account of the high switching force of the piezo actuator, a servo-valve flight time (time between end stops) of just 50 µs and a practically bounce-free switching time are achieved. This meets the contradictory requirements with regard to high nozzle-needle speeds and extremely small reproducible injected fuel quantities.

Similarly to the solenoid-valve injector, a control quantity is terminated via the servo valve to activate an injection. Necessitated by the design, however, the piezo inline

injector otherwise contains no leakage points between the high-pressure and low-pressure circuits. The result is an increase in the hydraulic efficiency of the overall system.

Up to eight injection pulses can be set per injection cycle. In this way, the injection can be adapted to the requirements of the respective engine operating point. The maximum possible number of injection pulses decreases in the upper rpm range.

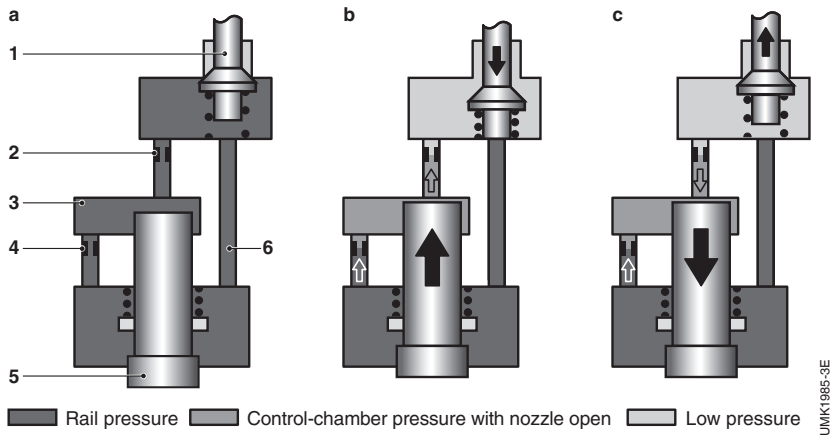
Operating principle of the servo valve

The nozzle needle on a piezo inline injector is controlled indirectly by the servo valve. The desired injected fuel quantity is set, taking into account the prevailing rail pressure, via the activation duration of the valve. In its non-triggered state, the injector is in the starting position and the servo valve is closed (Figure 9a), i.e., the high-pressure section is separated from the low-pressure section. The nozzle is kept closed by the rail pressure exerted in the control chamber.

When the piezo actuator is triggered, it increases in length and this increase is transmitted via the hydraulic coupler to the servo valve. This process opens the servo valve and closes the bypass passage (Figure 9b). The flow-rate ratio between the outlet and inlet restrictors lowers pressure in the control chamber and the nozzle opens. The control quantity flows via the servo valve to the low-pressure circuit of the overall system and from there back to the fuel tank.

To start the closing process, the actuator is discharged, and the servo valve closes to release the bypass passage at the same time (Figure 9c). The control chamber is then refilled by reversing the inlet and outlet restrictors, and the nozzle needle is closed again. The injection process is completed as soon as the nozzle needle reaches the nozzle seat again.

Figure 9: Operating principle of servo valve
a) Start position, b) Nozzle needle opens (bypass closed),
c) Nozzle needle closes (bypass open).
1 Valve pin, 2 Outlet restrictor, 3 Control chamber,
4 Inlet restrictor, 5 Nozzle needle, 6 Bypass.



Operating principle of the hydraulic coupler

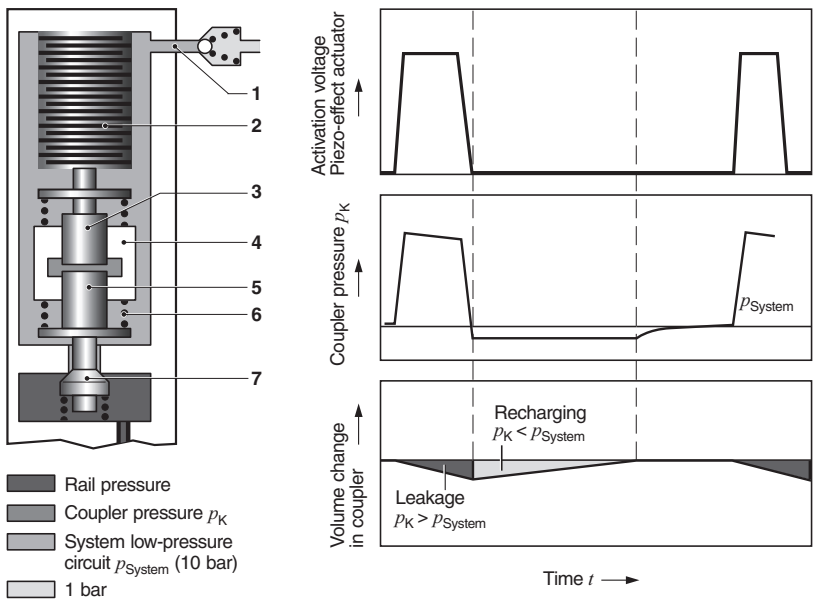
Another vital component in the piezo inline injector is the hydraulic coupler (Figure 10). It provides for compensation of length tolerances of the steel and ceramic components (e.g., caused by the varying thermal expansion of ceramic and steel or by clamping forces on the holder body). On the other hand, it adjusts the translation of actuator stroke and actuator force to the level required on the servo-valve side. The translation ratio is obtained from the diameters of the coupler plunger and valve plunger.

The actuator module and the hydraulic coupler are immersed in the diesel fuel flow, which through the system low-pressure circuit at the injector return is at a pressure of about 10 bar. When the actuator is not triggered, pressure in the hydraulic coupler is in equilibrium with its surroundings, and the coupler does not exert any force on the valve pin. Changes in length caused by temperature or clamping forces acting on the holder body are compensated by small leakages that flow via the guide clearances of the coupler and valve plungers between the coupler gap and the coupler surrounding area. This maintains a coupling of forces between the piezo actuator and the servo valve at all times.

A voltage dependent on the rail pressure (110 to 160 V) is applied to the actuator in order to create an injection event. This causes the pressure in the coupler to increase and a switching force to be exerted on the valve pin. The servo valve opens if this switching force exceeds the closing force on the valve pin caused by the rail pressure. On account of the higher

Figure 10: Operating principle of hydraulic coupler

1 Low-pressure rail with pressure-holding valve (fuel return), 2 Piezo actuator, 3 Coupler plunger, 4 Hydraulic coupler, 5 Valve plunger (lower coupler plunger), 6 Valve-plunger spring, 7 Valve pin.



pressure in the coupler than in the surroundings, a small leakage volume flows through the plunger guide clearances out of the coupler into the low-pressure circuit (10 bar) of the injector. Even when the coupler is repeatedly actuated in quick succession during an engine combustion cycle, there is no significant effect on the injector function when the coupler is drained.

The quantity missing in the hydraulic coupler is refilled in the pauses between the triggering pulses of the piezo actuator. This occurs in the reverse direction via the plunger guide clearance, whereby the valve-plunger spring generates a vacuum pressure of max. 10 bar inside the coupler against the surroundings. The guide clearances and the low-pressure level are matched to refill the hydraulic coupler entirely before the next engine combustion cycle starts.

Further development of piezo injectors

The further development of piezo injectors is being driven forward in two directions:

– higher maximum permissible injection pressures,
– shortening of the necessary idle times between the injections.

High-pressure pumps

Design and requirements

The high-pressure pump is the interface between the low-pressure and high-pressure stages of the common-rail system. Its function is to make sure there is always sufficient fuel under pressure in all engine operating conditions. At the same time, it must operate for the entire service life of the vehicle. This includes providing a fuel reserve that is required for quick engine starting and rapid pressure rise in the fuel rail.

The high-pressure pump generates a constant system pressure for the high-pressure accumulator (fuel rail) independent of fuel injection. For this reason, fuel – compared to conventional fuel-injection systems – is not compressed during the injection process.

3-, 2- and 1-plunger radial-piston pumps are used as the high-pressure pump to generate pressure. In a 3-plunger pump an eccentric shaft provides the stroke motion for the pump plungers, while in 2- and 1-plunger pumps this is provided by a camshaft. 2-plunger in-line fuel-injection pumps are also used in commercial vehicles.

Preferably, the high-pressure pump is fitted to the diesel engine at the same point as a conventional distributor injection pump. The pump is driven by the engine via coupling, gearwheels, chain, or toothed belt. Pump speed is therefore coupled to engine speed via a fixed gear ratio.

High-pressure pumps are used in a number of different designs in passenger cars and commercial vehicles. Within the pump generations there are versions that have different delivery rates (50 – 550 *l*/h) and delivery pressures (900 – 2,500 bar).

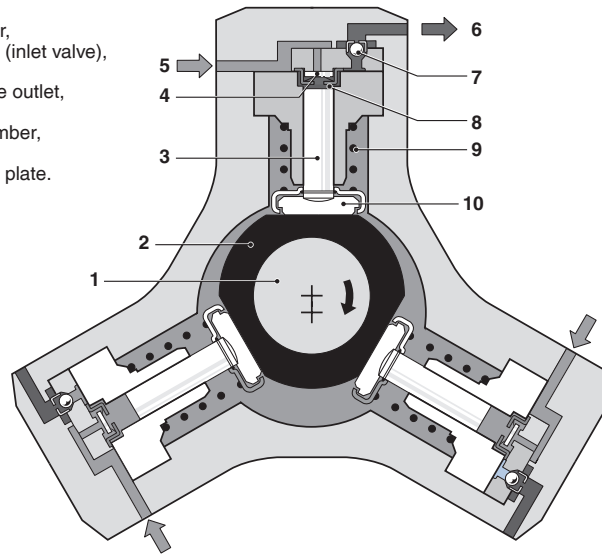
3-plunger radial-piston pump

Design

The drive shaft is centrally mounted in the housing of the high-pressure pump (Figures 11 and 12). The pump elements are arranged radially with respect to the

Figure 11: 3-plunger radial-piston pump (cross-section)

1 Drive shaft with eccentric,
2 Polygon ring,
3 Pump plunger,
4 Suction valve (inlet valve),
5 Fuel supply,
6 High-pressure outlet,
7 Outlet valve,
8 Element chamber,
9 Valve spring,
10 Plunger base plate.



central bearing and offset by 120°. The polygon ring fitted to the drive-shaft eccentric forces the pump plunger to move up and down. Force is transmitted between the eccentric shaft and the pump plunger via the plunger base plate attached to the plunger base.

Fuel delivery and compression

The presupply pump – an electric fuel pump or a mechanically driven gear pump – delivers fuel via a filter and water separator to the inlet of the high-pressure pump. The inlet is located inside the pump on passenger-car systems with a gear pump flanged to the high-pressure pump. An overflow valve is fitted behind the inlet. If the delivery pressure of the presupply pump exceeds the opening pressure (0.5 – 1.5 bar) of the overflow valve, the fuel is pressed through the restriction bore of the overflow valve into the lubrication and cooling circuit of the high-pressure pump. With its eccentric, the drive shaft moves the pump plunger up and down to mimic the eccentric lift. Fuel passes through the suction valve into the element chamber and the pump plunger moves downward (induction stroke).

When the bottom-dead center of the pump plunger is exceeded, the suction valve closes, and the fuel in the element chamber can no longer escape. It can then be pressurized beyond the delivery pressure of the presupply pump. When the rising pressure exceeds the back pressure from the rail, the outlet valve opens and the compressed fuel passes into the high-pressure circuit. The high-pressure ports of the three pump elements are combined inside the pump housing so that only one high-pressure line runs to the rail.

The pump plunger continues to deliver fuel until it reaches its top-dead center position (delivery stroke). The pressure then drops so that the outlet valve closes. The pump plunger moves downward in response to the force of the valve spring and the remaining fuel in the dead volume is depressurized.

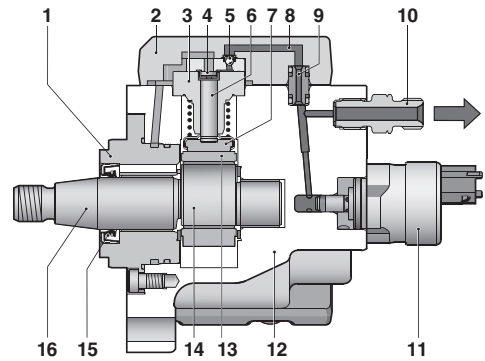
When the pressure in the element chamber drops below the difference between the pre-delivery pressure and the opening pressure of the suction valve, the suction valve reopens and the process starts over.

Transmission ratio

The delivery quantity of a high-pressure pump is proportional to its rotational speed. In turn, the pump speed is dependent on the engine speed. The transmission ratio between the engine and the pump is determined in the process of adapting the fuel-injection system to the engine so as to limit the volume of excess fuel delivered. At the same time, it makes sure that the engine's fuel de-

Figure 12: 3-plunger radial-piston pump (longitudinal section)

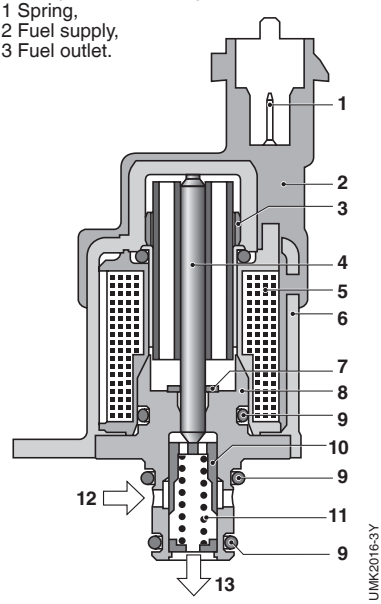
1 Flange,
2 Cylinder head,
3 Pump barrel,
4 Suction valve (inlet valve),
5 Outlet valve,
6 Pump plunger,
7 Plunger base plate,
8 High-pressure passage,
9 Connecting piece,
10 High-pressure connection,
11 Pressure-control valve (for constant-delivery high-pressure pump),
12 Pump housing,
13 Polygon ring,
14 Eccentric,
15 Shaft seal,
16 Input shaft.



mand at WOT is covered to the full extent. Possible transmission ratios relative to the crankshaft are between 1:2 and 5:6, i.e. the high-pressure pump is geared up. Pump speed is thus higher than engine speed. Higher transmission ratios are required for commercial vehicles on account of the low engine speeds.

Delivery rate
As the high-pressure pump is designed for high delivery quantities, there is a surplus of pressurized fuel when the engine is idling or running in part-load range. This excess fuel is returned in first-generation systems to the fuel tank via the pressure-control valve seated on the rail or flanged to the pump. As the compressed fuel expands, the energy imparted by compression is lost; overall efficiency drops. Compressing and then expanding the fuel also heats the fuel.

Figure 13: Design of metering unit
1 Plug with electrical interface,
2 Solenoid housing,
3 Bearing,
4 Armature with tappet,
5 Solenoid-valve winding with coil body,
6 Bowl,
7 Residual air-gap washer,
8 Magnetic core,
9 O-ring,
10 Plunger with metering slots,
11 Spring,
12 Fuel supply,
13 Fuel outlet.



Demand control
An improvement in energy efficiency is possible by means of fuel-delivery control of the high-pressure pump on the fuel-delivery side (suction side). Fuel flowing into the pump elements is metered by an infinitely variable solenoid valve (metering unit) mounted on the high-pressure pump (Figure 13). This valve adapts the fuel quantity delivered to the rail to system demand via the metering slots in the plunger. The plunger operated by the solenoid valve frees up a metering orifice depending on its position via the metering slots. The solenoid valve is activated by means of a PWM signal (pulse width modulation).

This fuel-delivery control not only drops the performance demand of the high-pressure pump, it also reduces the maximum fuel temperature.

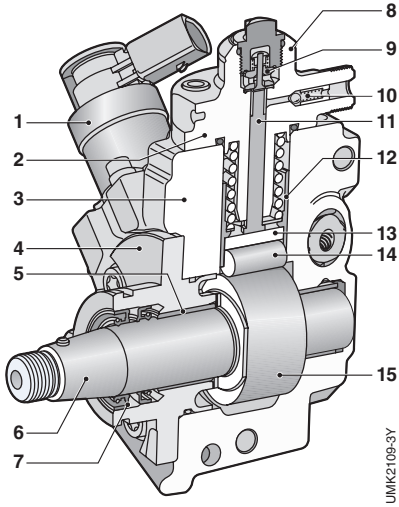
1- and 2-plunger radial-piston pumps

Requirements
The delivery stroke of the pump elements causes pressure pulsations in the rail which give rise to fluctuations in injected fuel quantity in the 3-plunger pumps. In order to comply with the ever more stringent emissions limits, precision of injection coupled with minimal fluctuations in injected fuel quantity is increasingly gaining in importance. 1- and 2-plunger radial-piston pumps facilitate injection-synchronous delivery, i.e. the delivery stroke of the pump elements is synchronous with the induction stroke of the engine cylinders. In this way, the pump delivery for each engine cylinder is always at the same crankshaft angle.

With one or two pump elements, all engines with three to eight cylinders can be injection-synchronously operated by adapting the transmission ratio of 1:2 to 1:1 between engine and pump speeds.

Design
This high-pressure pump is a radial-piston pump of 1- or 2-plunger design. Its components are (Figure 14):
– Aluminum housing, to which only low pressure is applied
– One or two pump elements with high-pressure-resistant steel cylinder heads with integrated high-pressure valve and high-pressure port
– Cam drive assembly with roller tappet, which converts the rotary motion of the camshaft via the cams (dual cam with 180° offset) into a stroke motion of the pump in the cylinder head. The camshaft is guided in the mounting flange and housing in two friction bearings.

Figure 14: 1-plunger radial-piston pump
1 Metering unit,
2 Pump element,
3 Pump housing,
4 Mounting flange,
5 Friction bearing,
6 Input shaft (camshaft),
7 Shaft seal,
8 Cylinder head,
9 Suction valve (inlet valve),
10 High-pressure valve (non-return valve) in high-pressure port (fuel inlet not visible in this illustration),
11 Pump plunger,
12 Roller tappet,
13 Roller support,
14 Drive roller,
15 Dual cam.



The high pressure is generated in the pump element. 1- or 2-plunger pumps are used, depending on the engine's displacement and number of cylinders. Two pump elements are required to cover the fuel demand of larger engines. In the 2-plunger version, the pump elements are arranged in a vee configuration at 90° to each other.

The great overlap length between the cylinder wall and the pump plunger results in low leakage losses when the fuel is compressed. On the other hand, short leakage times result – on account of the high delivery frequency (two strokes per rotation per plunger) and the small dead volume in the cylinder head – in further optimization of efficiency and thus reduced fuel consumption.

The 90° vee configuration of the cylinder heads in the 2-plunger pump means that there is no overlapping of the induction strokes. The charge of the two pump elements is therefore identical (uniformity of delivery).

The high-pressure port is connected to the rail by way of one (in the 1-plunger pump) or two (in the 2-plunger pump) high-pressure lines. The high pressure is not condensed in the housing, but instead routed directly outwards from the cylinder head. The housing therefore does not require any high-pressure- and strength-increasing measures.

Low-pressure circuit
The total fuel delivered by the presupply pump (electric fuel pump or gear pump flanged to the high-pressure pump) is routed through the pump interior to the overflow valve and to the metering unit. The volume of fuel used for lubrication and cooling is thus greater than in previous pumps. The overflow valve regulates the pump internal pressure and thereby protects the housing against overpressure.

The entire low-pressure path is dethrottled on account of the large cross-sections in such a way as to ensure that the pump elements are filled even at high engine speeds. Fuel-quantity metering is

performed on the low-pressure side with the metering unit or the electric suction valve (see Demand control).

High-pressure circuit
The fuel pilot-controlled by the metering unit passes in the induction phase through the suction valve into the element chamber, and is compressed to high pressure during the ensuing delivery phase and delivered by the high-pressure valve and the high-pressure line to the rail.

Demand control with metering unit
The concept of the metering unit used on these pumps corresponds to that of the metering unit used for the 3-plunger radial-piston pump (Figure 13); however, they differ in their design.

Demand control with electric suction valve
The electric suction valve is, like the metering unit, a solenoid valve which is electrically activated. Unlike the metering unit, which is a proportional valve, the electric suction valve is a switching valve.

The electric suction valve is an actively controlled valve (Figure 15) that is seated in place of a conventional suction valve on the injection-pump cylinder head. The plunger chamber is filled with fuel in the high-pressure-pump suction phase. The suction valve is initially kept open during the delivery phase. The delivered fuel is routed back into the low-pressure circuit until the desired fuel quantity is reached in the plunger chamber. The suction valve is then closed and the fuel quantity remaining in the plunger chamber is delivered to the rail. The electric suction valve is activated as a function of the cam-rotation angle.

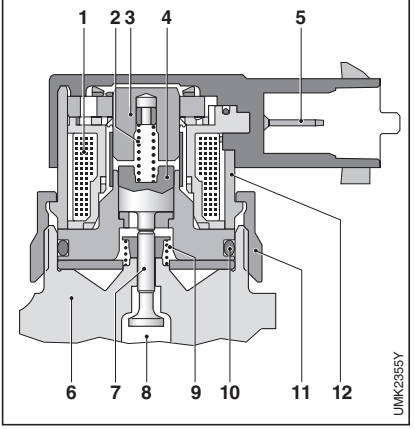
Design and function
The electric suction valve consists of the magnet group and the hydraulic module. The suction valve is open at zero current. In its non-activated state the suction valve is kept open via the armature spring against the hydraulic flow forces. When the suction valve is activated, the solenoid coil is energized to build up a magnetic field. The magnetic field generates a force in the armature which acts against the armature spring. If the magnetic force is greater than the spring force, the armature moves towards the pole body and thereby closes the suction valve.

Features of electric suction valve
The electric suction valve offers more flexibility in activation for high-efficiency engines. It is possible to determine with the cam-angle-dependent control process when the pump begins fuel delivery. This enables precise, dynamic rail-pressure control.

Active opening of the suction valve creates very slight throttling effects during the suction phase, which results in high hydraulic pump efficiency.

A further feature of the electric suction valve to be highlighted is its capacity not to deliver any fuel to the rail despite the pump being driven. This results in the possibility of engine cutout.

Figure 15: Electric suction valve
1 Solenoid coil,
2 Armature spring,
3 Pole body,
4 Armature,
5 Flat connector,
6 Cylinder head of high-pressure pump,
7 Valve plunger,
8 Plunger chamber of high-pressure pump,
9 Compression spring,
10 O-ring,
11 Union nut,
12 Solenoid sleeve.



2-plunger in-line piston pump

Design

This demand-controlled high-pressure pump for rail pressures up to 2,500 bar is only used in commercial vehicles. This is a 2-plunger pump with an in-line design, i.e. the two pump elements are arranged one after the other in relation to the camshaft axial direction (Figure 16). This high-pressure pump comes in both oil-lubricated and fuel-lubricated variants.

A spring seat forms the positive link between the roller tappet and the pump plunger. The rotary motion of the camshaft is converted via the cams into a stroke motion of the pump plungers. The plunger spring ensures pump-plunger return. The combined inlet and outlet valve is located at the top of the pump element.

A gear presupply pump with a high gear ratio is located on the camshaft extension. Its function is to draw fuel via the fuel inlet from the fuel tank and route it through the fuel outlet to the fuel fine filter. From there, the fuel passes through another line to the metering unit located on the upper section of the high-pressure pump.

Lube oil is supplied either directly via the mounting flange of the pump or via a side-mounted inlet. The lube oil is returned to the engine oil pan.

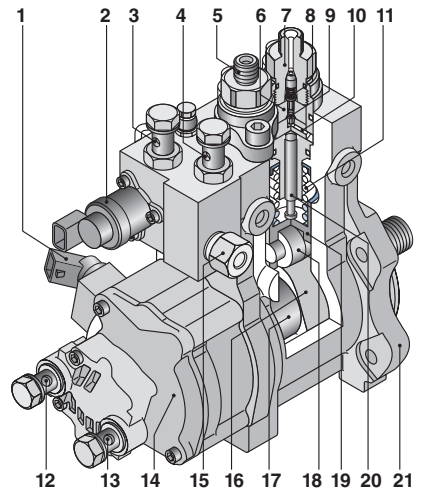
Operating principle

When the pump plunger moves from top dead center towards bottom dead center, the suction valve opens on account of the fuel pressure (pre-delivery pressure). The fuel is drawn into the element chamber in response to the downward movement of the pump plunger. The outlet valve is closed by the valve spring.

As the pump plunger moves upwards, the suction valve closes and the enclosed fuel is compressed. When the rail pressure is exceeded, the outlet valve opens and the fuel is delivered through the high-pressure port to the rail. This increases the pressure in the rail. The rail-pressure sensor measures the pressure, from which the engine ECU calculates the triggering signals (PWM) for the metering unit. The metering unit regulates the fuel quantity provided for compression according to the current demand.

Figure 16: 2-plunger in-line piston pump

1 Speed sensor (pump speed),
2 Metering unit,
3 Fuel supply for metering unit (from fuel filter),
4 Fuel return line to fuel tank,
5 High-pressure port,
6 Valve body,
7 Valve holder,
8 Outlet valve with valve spring,
9 Suction valve (inlet valve) with valve spring,
10 Fuel supply to pump element,
11 Plunger spring,
12 Fuel supply from fuel tank,
13 Fuel outlet to fuel filter,
14 Gear presupply pump,
15 Overflow valve,
16 Concave cam,
17 Camshaft,
18 Roller bolt with roller,
19 Roller tappet,
20 Pump plunger,
21 Mounting flange.



Rail

Function

The function of the rail is to maintain the fuel at high pressure. In so doing, the accumulator volume dampens pressure fluctuations caused by fuel pulses delivered by the pump and the fuel-injection cycles. This ensures that, when an injector opens, the injection pressure remains constant. On the one hand, the accumulator volume must be large enough to meet this requirement. On the other hand, it must be small enough to ensure a fast enough pressure rise on engine start.

Besides acting as a fuel accumulator, the fuel rail also distributes fuel to the injectors.

Application

The tubular rail can vary depending on the various constraints which determine the way it is fitted to the engine. It has connections for fitting the rail-pressure sensor and the pressure limiter or the pressure-control valve (Figure 17).

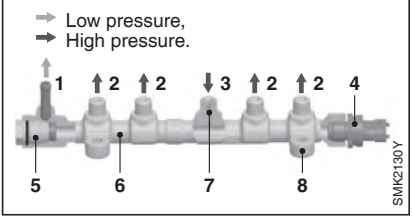
The pressurized fuel delivered by the high-pressure pump passes via one or two high-pressure fuel lines to the rail inlet. From there, it is distributed via high-pressure lines to the individual injectors. The cavity inside the fuel rail is permanently filled with pressurized fuel during engine operation.

The fuel pressure is controlled by means of Electronic Diesel Control (EDC), whereby the fuel pressure is measured by the rail-pressure sensor and – depending on the system – regulated by demand control or via the pressure-control valve to the desired value. The pressure limiter is used as an alternative to the pressure-control valve – depending on system requirements – and its function is to limit fuel pressure in the fuel rail to the maximum permissible pressure in the event of a fault.

Some rail types make use of restrictors in the inlet and outlet which additionally dampen pressure pulsations caused by pump delivery and injections. If fuel is now drawn from the rail for an injection, the pressure is the rail remains virtually constant.

Figure 17: Design of a rail with pressure limiter

1 Fuel return (low pressure),
2 High-pressure ports injectors,
3 High-pressure port to high-pressure pump (one or two ports),
4 Rail-pressure sensor.
5 Pressure limiter,
6 Rail body,
7 Restrictor (pressed-in, optional),
8 Mounting lug (engine mounting).



Rail types

Two different rail types are used: the forged rail (hot-forged rail) and the welded rail (laser-welded rail). Both types are used by Bosch, the hot-forged rail being the preferred variant. The welded rail is used in series production up to 2,000 bar and is not subject to further development.

In the case of a hot-forged rail, the blank for machining is produced by forging from bar stock. The internal geometry and the rail interfaces of the rail body are produced by deep-hole boring, drilling and milling. A corrosion-resistant surface is then applied. Finally, the add-on components are mounted and functionally tested.

The forging process affords more possibilities in the shaping of the external geometry than laser welding. One advantage is the possibility of shaping the external geometry with regard to weight optimization.

The forged rail is used for passenger-car and commercial-vehicle applications in series production up to 2,500 bar, and a further pressure increase is planned for the next generations.

Time-controlled single-cylinder pump systems

Unit-injector system for passenger cars

System requirements

The electronic controlled unit injector is a single-cylinder pump injection system with integrated high-pressure pump and injection nozzle (Figure 1). The high-pressure line required in other injection systems between the injection pump and the injection nozzle is omitted. This means that the system has particularly good hydraulic performance.

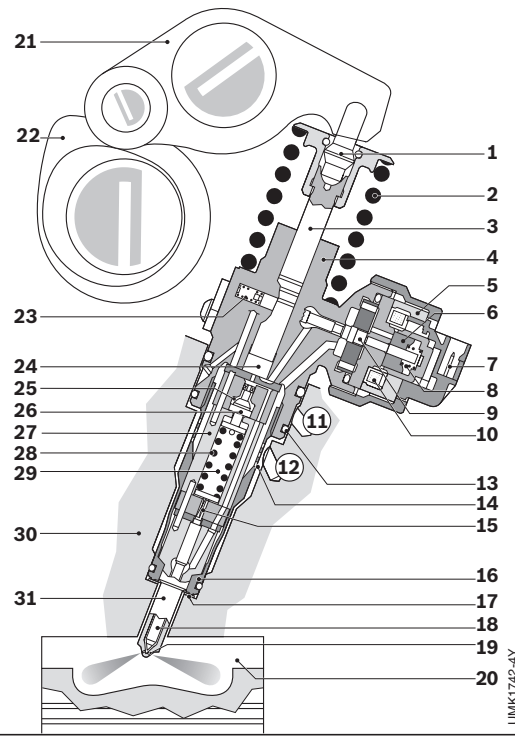
The unit injector is installed in the cylinder head between the valves, with the nozzle protruding into the combustion chamber. It is operated by rocker arms driven by the overhead engine camshaft. Each engine cylinder is allocated a separate unit injector. The start of injection and injection duration are calculated by an electronic control unit and controlled by the high-pressure solenoid valve mounted externally on the pump body.

The unit-injector system (UIS) is designed to meet the demands of modern direct-injection diesel engines with high levels of power density. It is characterized by its compact design, high injection pressures of up to 2,200 bar at full load, and mechanical-hydraulic pilot injection throughout the entire program-map range to substantially reduce combustion noise. The compact design enables a very low high pressure volume with correspondingly high hydraulic efficiency.

Unit-injector systems are no longer used for new developments.

Figure 1: Design of unit injector for passenger cars

1 Ball pin, 2 Return spring,
3 Pump plunger, 4 Pump body,
5 Magnetic core of high-pressure solenoid valve, 6 Armature,
7 Electrical connection,
8 Compensating spring,
9 Solenoid-valve needle,
10 Electromagnet coil,
11 Fuel return (low-pressure stage),
12 Fuel feed, 13 Seal,
14 Fuel inlet passages (approx. 350 laser-drilled holes as filter),
15 Hydraulic stop (damping unit),
16 Retaining nut,
17 Sealing washer,
18 Nozzle needle,
19 Nozzle-needle seat,
20 Engine combustion chamber,
21 Roller rocker arm,
22 Engine camshaft with actuating cam,
23 Solenoid-valve spring,
24 High-pressure chamber (element chamber),
25 Accumulator chamber,
26 Accumulator plunger (bypass plunger),
27 Spring retainer,
28 Nozzle spring (compression spring),
29 Spring-retainer chamber,
30 Engine cylinder head,
31 Integrated injection nozzle.



Operating principle

Intake stroke

The unit injector is filled during the intake stroke of the upward-moving pump plunger. Fuel flows from the low-pressure stage of the fuel supply through the inlet holes into the unit injector. The fuel enters the high-pressure chamber when the solenoid-valve seat is opened.

Prestroke

The rotation of the actuating cam causes the pump plunger to move downwards. When the solenoid valve is still open, the fuel is forced back by the pump plunger from the high-pressure chamber into the low-pressure stage of the fuel supply. Heat is also dissipated from the unit injector (cooling) with the fuel as it flows back.

Delivery stroke and pilot injection

The electronic control unit energizes the coil of the solenoid valve at a particular point in time so that the solenoid-valve needle is pressed onto the solenoid-valve seat and the connection between high-pressure chamber and low-pressure stage is closed (beginning of injection period, BIP). The downward movement of the pump plunger increases the fuel pressure in the high-pressure chamber. The nozzle-opening pressure for pilot injection is approximately 180 bar. When this pressure is reached, the nozzle needle is raised and pilot injection begins. In this phase the nozzle-needle lift is hydraulically limited by a damping unit located between nozzle needle and nozzle spring so that the low injected fuel quantity required can be accurately metered. The accumulator plunger remains on its seat for the time being because the nozzle needle opens first due to its greater hydraulically effective area on which the pressure acts.

As the pressure increases, the accumulator plunger (bypass plunger) is forced downwards and now lifts off its seat. A connection is established between high-pressure chamber and accumulator chamber. The resulting pressure drop in the high-pressure chamber, the increased pressure in the accumulator chamber and the simultaneous increase in the pre-tension of the nozzle spring cause the nozzle needle to close. Pilot injection is terminated. However, the accumulator plunger does not return to its starting position because in its open state it offers the fuel pressure a greater working area than the nozzle needle.

Main injection

The continuing movement of the pump plunger causes the pressure in the high-pressure chamber to increase further. The nozzle-opening pressure for main injection is at approximately 300 bar higher than pilot injection. There are two reasons for this. Firstly, the excursion of the accumulator plunger increases the pre-tension of the nozzle spring. Secondly, due to the bypassing of the accumulator plunger fuel must be forced from the spring-retainer chamber via a restrictor into the low-pressure stage of the fuel supply so that the fuel in the spring-retainer chamber is compressed to a greater extent (pressure-backing). The pressure-backing level is derived from the size of the restrictor in the spring retainer and can be varied. It is therefore possible to achieve a sensible compromise between a low opening pressure for pilot injection (for noise reasons) and as high an opening pressure as possible for main injection, especially at part load (reducing emissions). The stroke and shaft diameter of the bypass plunger determine the length of the interval, the so-called injection interval, between the end of pilot injection and the start of main injection.

When the nozzle-opening pressure is reached, the nozzle needle is raised and fuel is injected into the combustion chamber (actual start of injection). The pressure during the entire injection process continues to rise due to the high delivery rate of the pump plunger.

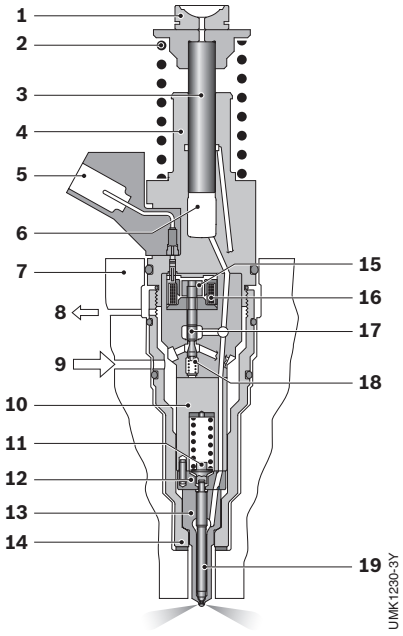
The current flow through the solenoid-valve coil is deactivated to terminate main injection. The solenoid valve opens and releases the connection between high-pressure chamber and low-pressure stage. The pressure collapses. When the nozzle-closing pressure is undershot, the injection nozzle closes and the injection process ends. At this point the accumulator plunger also returns to its starting position. The remaining fuel is returned to the low-pressure stage during the continuing

downward movement of the pump plunger (residual stroke). Heat is also dissipated from the unit injector in the process.

Electronic control allows the values for the start of injection and the injected fuel quantity to be selected as desired from those stored in the program map. This feature, together with the high injection pressures, makes it possible to achieve high power densities combined with low emission levels and low fuel consumption.

Figure 2: Design of unit injector for commercial vehicles

1 Sliding disk, 2 Return spring,
3 Pump plunger, 4 Pump body,
5 Electrical connection,
6 High-pressure chamber (element chamber),
7 Engine cylinder head,
8 Fuel return (low-pressure stage),
9 Fuel feed, 10 Spring retainer,
11 Pressure pin, 12 Intermediate ring,
13 Integrated injection nozzle,
14 Retaining nut, 15 Armature,
16 Electromagnet coil,
17 Solenoid-valve needle,
18 Solenoid-valve spring,
19 Nozzle needle.



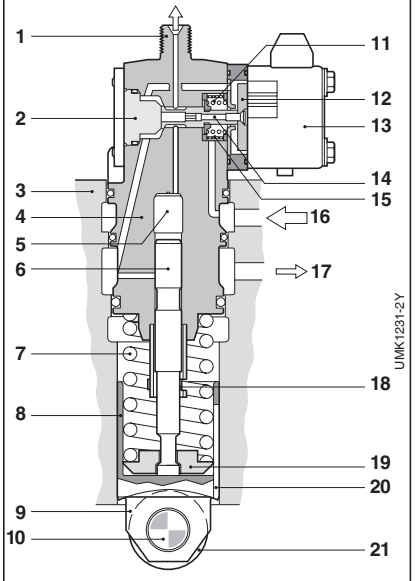
Unit-injector system for commercial vehicles

The unit-injector system for commercial vehicles is largely the same as the system for passenger cars. Due to the larger dimensions for the commercial-vehicle system, the solenoid valve can be integrated in the unit-injector here (Figure 2).

With regard to main injection, the commercial-vehicle system has the same operating principle as the passenger-car system. The systems differ with regard to pilot injection. In the unit-injector System

Figure 3: Design of unit pump

1 Hydraulic high-pressure port,
2 Solenoid-valve needle-travel stop,
3 Engine block, 4 Pump body,
5 High-pressure chamber,
6 Pump plunger, 7 Tappet spring,
8 Roller-tappet shell,
9 Roller tappet, 10 Roller-tappet pin,
11 Solenoid-valve spring,
12 Armature plate,
13 Solenoid-valve housing,
14 Solenoid-valve needle, 15 Filter,
16 Fuel feed, 17 Fuel return,
18 Pump-plunger retainer,
19 Spring seat, 20 Locating groove,
21 Tappet roller.



for commercial vehicles this can be electronically controlled in the lower speed and load range. This significantly reduces combustion noise and improves cold-starting performance.

Unit-pump system for commercial vehicles

The unit-pump system (UPS) is also a modular, time-controlled single-cylinder pump injection system, and is closely related to the unit-injector system. It is used in commercial-vehicle engines and in large engines. Each of the engine's cylinders is supplied by a separate module consisting of a high-pressure plug-in pump with integral, high-speed solenoid valve, a short high-pressure line, and an injection nozzle (Figures 3 and 4). This system is therefore also called Pump-Line-Nozzle (PLN). It permits injection pressures of up to 2,100 bar.

The separation of high-pressure generation and injection permits a simpler mounting on the engine. The hydraulic performance is nevertheless very good due to the shortest line possible. The unit pump is mounted on the side of the engine block. It is operated directly by an injection cam on the engine's camshaft via a roller tappet. The injection nozzle is installed with a nozzle holder in the cylinder head. The lines are made of high-strength, seamless steel tubes. They must be of equal length for the individual pumps of an engine.

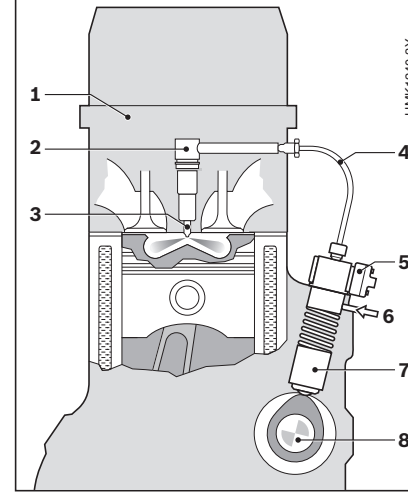
The method of solenoid-valve actuation is the same as that of the unit-injector system. When the solenoid valve is open, the fuel can be drawn into the pump barrel during the pump plunger's intake stroke, and return during the delivery stroke. Only when the solenoid valve is energized, and thus closed, can pressure build up in the high-pressure system between the pump plunger and the nozzle during the pump-plunger delivery stroke. Fuel is injected into the combustion chamber of the engine once the nozzle-opening pressure is exceeded.

Electronic control

The solenoid valves are actuated by an electronic control unit. The ECU analyzes all of the relevant status parameters in the system relative to the engine and its environment, and defines the start of injection and injected fuel quantity for the operating state of the engine at any given time. The start of injection is also controlled by a BIP signal (beginning of injection period) in order to balance out the tolerances in the overall system. Start of injection is synchronized with engine piston position by analysis of the signals from an incremental trigger wheel.

In addition to the basic fuel-injection functions, there are further functions for improving driving smoothness (e.g. surge dampers, idle-speed governors, adaptive cylinder equalization). Diagnosis of the fuel-injection system and the engine is also included in the range of functions. The ECU communicates with other electronic vehicle components (e.g. antilock braking system (ABS), the traction control system (TCS) or transmission-shift control system) via a CAN data bus.

Figure 4: Unit-pump system
1 Engine, 2 Nozzle holder,
3 Injection nozzle, 4 High-pressure line,
5 Solenoid valve, 6 Fuel feed,
7 High-pressure pump (unit pump),
8 Camshaft.



Diesel distributor injection pumps

Diesel distributor injection pumps were used in large numbers from 1962 until after 2000. These pumps are used in 3-, 4-, 5- and 6-cylinder diesel engines in passenger cars, tractors and light and medium commercial vehicles. They generate up to 50 kW per cylinder, depending on engine speed and combustion system. Distributor injection pumps for direct-injection (DI) engines achieve a peak injection pressure of up to 1,950 bar in the nozzle at engine speeds up to 4,500 rpm.

A distinction is made between distributor injection pumps with mechanical control, and those with an electronic governor available in port-controlled versions with a rotary-magnet actuator and versions with solenoid-valve open-loop control.

In passenger cars and commercial vehicles, distributor injection pumps have been superseded by common-rail systems.

Axial-piston distributor pumps

Body

Fuel-supply pump

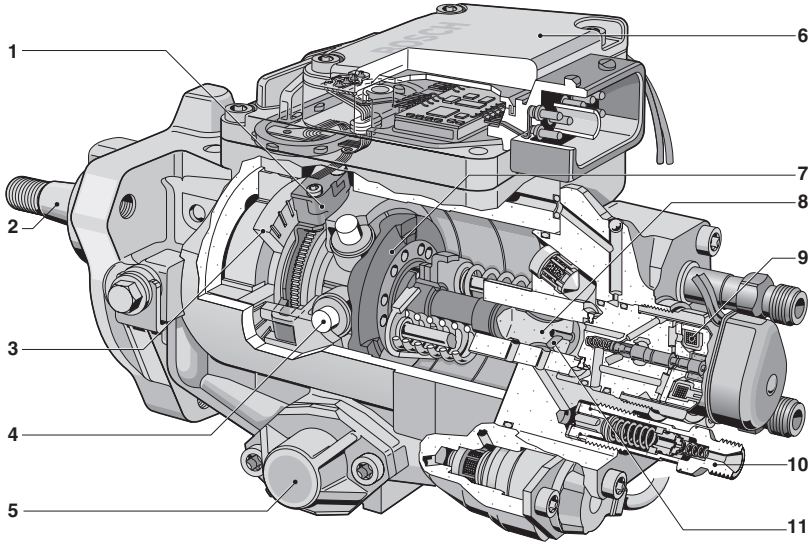
If no presupply pump is fitted in the fuel-injection system, this integral vane-type supply pump draws fuel from the tank and, together with a pressure-control valve, generates an internal pump pressure which increases with engine speed.

High-pressure pump

The axial-piston distributor pump (VE pump) incorporates only one pump element for all cylinders. The distributor plunger displaces the fuel during its axial stroke and rotates at the same time to distribute the fuel through passages to the individual high-pressure connections (Figure 1).

Figure 1: Solenoid-valve-controlled axial-piston distributor pump

1 Angle-of-rotation sensor, 2 Drive shaft, 3 Support ring of vane-type supply pump, 4 Roller ring, 5 Timing device, 6 Pump control unit, 7 Cam plate, 8 Distributor plunger, 9 High-pressure solenoid valve, 10 High-pressure connection, 11 Distributor bore.



A clutch unit transfers the rotation of the drive shaft to the cam plate and the distributor plunger fixed to it. When doing so, the claws of the drive shaft and the cam plate grip into the yoke positioned in between. The cam lifts on the underside of the cam plate turn against the rollers of the roller ring. As a result, the cam plate and distributor plunger perform a stroke movement in addition to the rotation. During each rotation of the drive shaft, the distributor plunger completes a number of strokes equal to the number of engine cylinders to be supplied.

The pump delivers fuel for as long as the spill port in the distributor plunger remains closed off during the working stroke. Delivery ends when the spill port is uncovered by the control collar (Figure 2).

Electronic control for distributor injection pumps with rotary-magnet actuator

In contrast to the mechanically controlled distributor injection pump, the pump with rotary-magnet actuator has an electronic governor and an electronically controlled timing device (Figure 2).

Electronic governor

An eccentrically mounted ball head provides the connection between the distributor injection pump control collar and the rotary-magnet actuator. The actuator's rotary setting determines the position of the control collar, and with it the effective stroke of the pump. A non-contacting position sensor is connected to the rotary-magnet actuator.

The control unit receives various signals from the sensors: accelerator-pedal position; engine speed; air, coolant and fuel temperature; charge-air pressure; atmospheric pressure; etc. It uses these input variables to determine the correct injected fuel quantity, which is then converted to a specific control-collar position with the aid of program maps stored in the unit's memory. The control unit varies the excitation current to the rotary-magnet actuator until it receives a signal indicating convergence between the set-point and actual values for control-collar position.

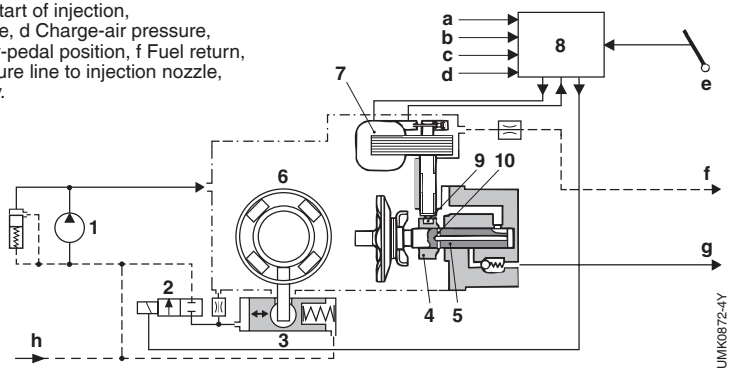
Electronically controlled timing device

The hydraulic timing device with the timing-device solenoid valve rotates the roller ring, depending on the load state and engine speed in such a way that the

Figure 2: Electronic control for axial-piston distributor pumps with rotary-magnet actuator

1 Vane-type supply pump, 2 Timing-device solenoid valve, 3 Timing-device plunger,
4 Control collar, 5 Distributor plunger, 6 Roller ring (view turned through 90°),
7 Rotary-magnet actuator with position sensor, 8 ECU,
9 Eccentrically mounted ball head, 10 Spill port.

Input and output variables:
a Speed, b Start of injection,
c Temperature, d Charge-air pressure,
e Accelerator-pedal position, f Fuel return,
g High-pressure line to injection nozzle,
h Fuel supply.



start of delivery – in relation to the engine piston position – is advanced or retarded. The signal from a sensor in the nozzle-holder assembly, which indicates when the nozzle begins to open, is compared with a programmed setpoint value. The timing-device solenoid valve connected to the working chamber of the plunger in the timing device varies the pressure above the timing-device plunger and thus the position of the timing device. The actuation clock ratio of the timing-device solenoid valve is varied until the setpoint and actual values agree.

Electronic control for solenoid-valve-controlled distributor injection pumps

In the case of solenoid-valve-controlled distributor injection pumps (Figure 1), the fuel is metered by a high-pressure solenoid valve which directly closes off the pump's element chamber. This permits even greater flexibility in fuel metering and for varying the start of injection. The main assemblies are the high-pressure solenoid valve, the ECU, and the incremental angle/time system for angle/time control of the solenoid valve using an angle-of-rotation sensor integrated in the pump.

The solenoid valve closes to define the start of delivery, which then continues until the valve opens. The injected fuel quantity is determined by the length of time the valve remains closed. Solenoid-valve control permits rapid opening and closing of the element chamber independently of engine speed. In contrast to mechanically governed pumps and pumps with a rotary-magnet actuator, direct triggering by means of solenoid valves results in lower dead volumes, improved high-pressure sealing, and therefore greater efficiency.

The fuel-injection pump is equipped with its own, integral control unit for precise start-of-delivery control and fuel metering. Individual pump program maps and example-specific calibration data are stored in this ECU.

The engine control unit determines the start of injection and delivery on the basis of engine operating parameters, and sends this data to the pump control unit via the data bus. The system can control both the start of injection and the start of delivery.

The pump control unit also receives the injected fuel quantity signal via the data bus. This signal is generated by the engine control unit according to the accelerator-pedal signal and other parameters for torque demand. In the pump control unit, the injected fuel quantity signal and the pump speed for a given start of delivery are taken as the input variables for the pump map on which the corresponding actuation period is stored as degrees of cam rotation.

And finally, the actuation of the high-pressure solenoid valve and the desired period of actuation are determined on the basis of the angle-of-rotation sensor integrated in the distributor injection pump. This sensor is used for angle/time control. It consists of a magnetoresistive sensor and a reluctor ring divided into 3° increments interrupted by a reference mark for each cylinder. The sensor determines the precise angle of cam rotation at which the solenoid valve opens and closes. This requires the pump control unit to convert timing data to angular position data and vice versa.

The low fuel-delivery rates at the start of injection, which result from the design of the distributor injection pump, are further reduced by the use of a two-spring nozzle holder. With a hot engine, these low delivery rates permit low basic noise levels.

Pilot injection

Pilot injection allows the combustion noise to be further reduced without sacrificing the objects of the system's design which aim at generating maximum power output at the rated-power operating point. Pilot injection does not require additional hardware. Within a matter of milliseconds, the electronic control unit actuates the solenoid valve twice in rapid succession. For the first injection, only a small amount of fuel is discharged to condition the combustion chamber. The solenoid valve controls the injected fuel quantity with a high degree of precision and dynamic response. Typical pilot-injection fuel quantities are 1.5 mm^3.

Radial-piston distributor pumps

Body

High-pressure pump

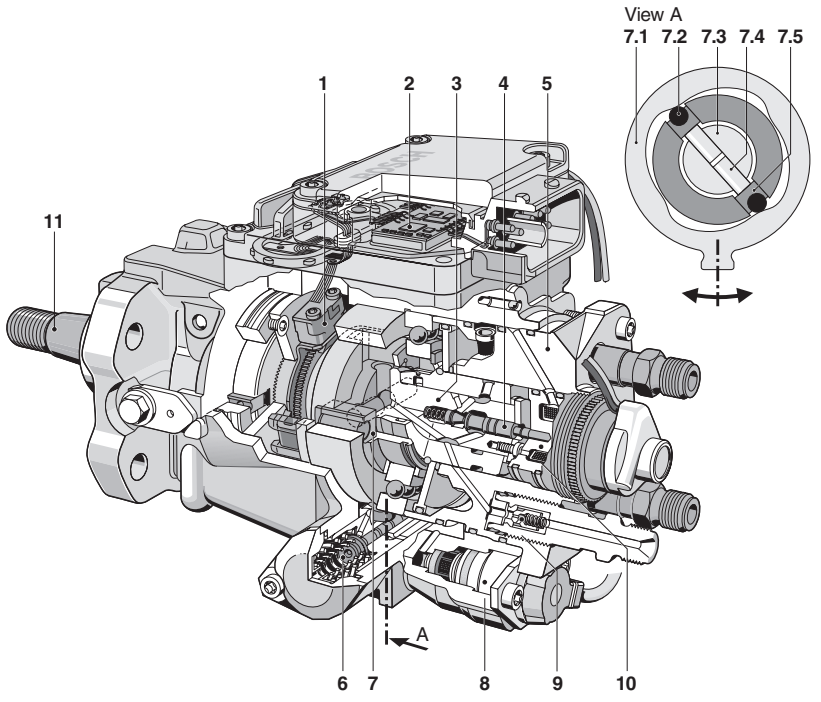
The radial-piston high-pressure pump (VR pump, Figure 3) is driven directly by the distributor-pump drive shaft. It comprises the cam ring, the roller supports and rollers, the delivery plunger, the drive plate, and the front section (head) of the distributor shaft.

The drive shaft drives the drive plate by means of radially positioned guide slots. The guide slots simultaneously act as the locating slots for the roller supports. The roller supports and the rollers held by them run around the inner cam profile of the cam ring that surrounds the drive shaft. The number of cams corresponds to the number of cylinders in the engine.

The drive plate drives the distributor shaft. The head of the distributor shaft holds the delivery plungers which are aligned radially to the drive-shaft axis (hence the name “radial-piston distributor pump”).

Figure 3: Solenoid-valve-controlled radial-piston distributor pump

1 Sensor (angle/time system), 2 Pump control unit, 3 Distributor shaft, 4 Solenoid-valve needle, 5 Valve body, 6 Timing device, 7 Radial-piston pump, 7.1 Cam ring, 7.2 Roller, 7.3 Distributor shaft, 7.4 Delivery plunger, 7.5 Roller support, 8 Timing-device solenoid valve, 9 Delivery valve, 10 High-pressure solenoid valve, 11 Drive shaft.



The delivery plungers rest against the roller supports. As the roller supports are forced outwards by centrifugal forces, the deliver plungers follow the profile of the cam ring and describe a cyclical-reciprocating motion.

When the delivery plungers are pushed inwards by the cams, the volume in the central plunger chamber between the delivery plungers is reduced. This compresses and pumps the fuel when the solenoid valve is closed. The fuel is directed through passages in the distributor shaft at defined times to the appropriate outlet delivery valves.

As the cam-drive design employs a direct, positive link, flexibility and compliance remain minimal, so the performance potential is greater. Fuel delivery is shared between at least two radial plungers. The small inertial forces involved mean that steep (fast) cam profiles are possible. The fuel-delivery rate can be further increased by increasing the number of delivery plungers.

Radial-piston distributor pumps for direct-injection engines achieve element-chamber pressures up to 1,100 bar and pressures in the nozzle up to 1,950 bar.

Electronic control system

High-pressure solenoid valve

The high-pressure solenoid valve opens and closes in response to the triggering signals of the pump control unit. The length of time it remains closed determines the delivery period of the high-pressure pump. This means that the fuel quantity can be very precisely metered for each individual cylinder.

The high-pressure solenoid valve is controlled by regulating the current. The pump control unit identifies the contact of the valve needle in the valve seat using the current curve. This allows the exact point at which fuel delivery starts to be calculated and the start of injection to be controlled very precisely.

Timing devices

The hydraulically assisted timing device rotates the cam ring in such a way that the start of delivery – in relation to the engine piston position – is advanced or retarded. The interaction between the high-pressure solenoid valve and the timing device thus varies the start of injection and the injection pattern to suit the operating status of the engine.

The cam ring engages in a cross-slot in the timing-device plunger by means of an adjuster lug so that the axial movement of the timing-device plunger causes the cam ring to rotate. In the center of the timing-device plunger is a control collar which opens and closes the control ports in the timing-device plunger. In axial alignment there is a spring-loaded hydraulic control plunger which defines the required position for the control collar. Under the control of the pump control unit, the timing-device solenoid valve modulates the pressure acting on the control plunger.

This hydraulically assisted timing device can apply higher displacement forces than the hydraulic timing device of the axial-piston distributor pump.

The timing-device solenoid valve acts as a variable throttle. It can vary continuously the control pressure so that the control plunger can assume any position between the fully advanced and fully retarded positions.

Version with full electronic circuitry on injection pump

The latest generation of distributor injection pumps are compact, self-contained systems incorporating an electronic control unit to control both the pump and the engine-management functions. As a separate engine control unit is no longer required, the fuel-injection system requires fewer connectors and the wiring harness is less complex, thus making installation simpler.

Fuel-injection system

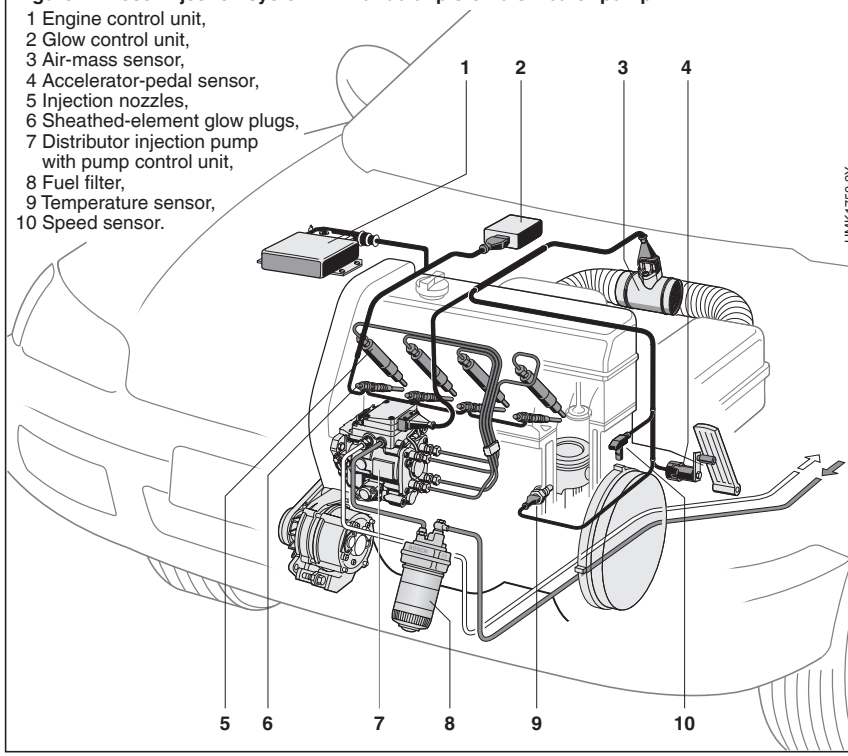
The injection pump is a component part of the fuel-injection system (Figure 4). A diesel fuel-injection system comprises the fuel-supply system (low-pressure stage), the high-pressure component, the injection components and the control system. The fuel-supply system accumulates and filters the fuel. If necessary, an additional fuel pump is present. The injection pump with the lines is the high-pressure component. It generates the high pressure and distributes it at the right time to the corresponding engine cylinders.

In distributor-pump injection systems, the injection component comprises the injection nozzles and the nozzle holders. These nozzle-holder assemblies come in a wide range of types. One nozzle-holder assembly is used on each cylinder. They are mounted with brackets or with hollow screws in the cylinder head. The functions of the injection nozzles are to inject metered quantities of fuel and to prepare the fuel, to shape the injection pattern, and to provide a seal against the combustion chamber. Each injection nozzle consists of the nozzle body with several injection orifices (up to 0.12 mm) and the nozzle needle. The needle is guided in the nozzle-body guide bore to ensure correct alignment between the injection orifices (which are at different angles in the nozzle body) and the engine combustion chamber.

The mechanical or electronic control system for the distributor injection pump is mounted on the pump itself. Some systems have a separate engine control unit. Versions with electronic control are equipped with numerous sensors and setpoint generators.

Figure 4: Diesel injection system with a radial-piston distributor pump

1 Engine control unit,
2 Glow control unit,
3 Air-mass sensor,
4 Accelerator-pedal sensor,
5 Injection nozzles,
6 Sheathed-element glow plugs,
7 Distributor injection pump with pump control unit,
8 Fuel filter,
9 Temperature sensor,
10 Speed sensor.



Start-assist systems for diesel engines

Preheating systems for passenger cars and light commercial vehicles

Hot diesel engines start spontaneously at low outside temperatures. Here, the spontaneous ignition temperature for diesel fuel of 250 °C is achieved with the engine turning at starting speed by compression. Cold diesel engines require, depending on their compression ratio, start assistance with glow plugs at low ambient temperatures (Figure 2).

The use of preheating systems extends far beyond pure start assistance. Thus, in driving mode the glow plugs can be consistently activated when the engine is hot in order to optimize combustion in all operating states.

Design and function of a preheating system

Preheating systems (Figure 1) consist essentially of glow plugs (GLP), the glow control unit (GCU), and preheating software in the engine management. The low-voltage preheating systems mostly used today require glow plugs with nominal voltages below the vehicle system voltage whose heating output is adapted via the electronic glow control unit to the engine requirement.

In DI engines the glow plug extends into the main combustion chamber of the engine cylinder. The air/fuel mixture is directed past the hot tip of the glow plug. In combination with the heating of the intake air during the compression cycle,

Figure 2: Startability of a diesel engine (without preheating) as a function of compression ratio ε
1 ε = 18,
2 ε = 16,
3 ε = 14.

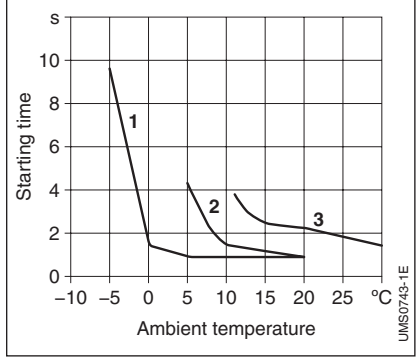
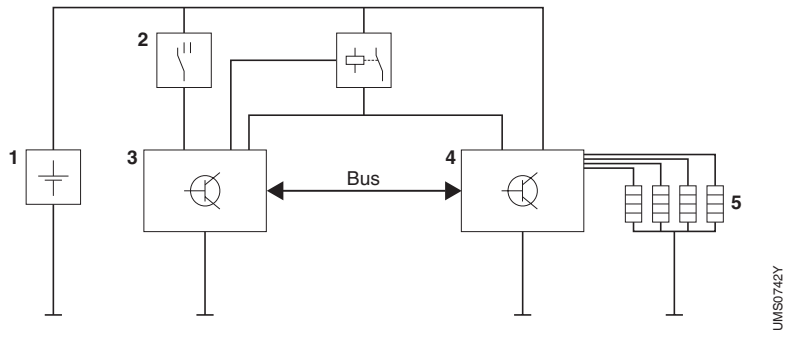


Figure 1: Preheating system
1 Battery, 2 Ignition/starting switch, 3 Engine ECU,
4 Glow control unit, 5 Glow plugs.



the ignition temperature is reached and the air/fuel mixture ignites.

The preheating software in the engine ECU starts and ends the preheating process as a function of the parameters stored in the software (e.g., ignition/starting switch on/off, coolant temperature). The glow control unit activates the glow plugs as directed by the engine ECU with vehicle system voltage via an electronic switch (transistor) during the preheating phases (pre-, standby, start, and post-heating). The supply voltage to the glow plugs is set with pulse-width-modulated activation to values below the vehicle system voltage. Direct operation at vehicle system voltage would destroy the glow element on account of the lower nominal voltage.

Preheating phases
- During preheating the glow plug is heated to operating temperature.
- During standby heating the preheating system maintains a temperature required for starting for a defined period.
- Start heating is applied while the engine is being revved up.
- The post-heating phase begins after starter release.
- Intermediate heating is activated after engine cooling due to coasting or to support particulate-filter regeneration.

In order to achieve as quickly as possible the heating temperature required for engine starting during pre-heating, the glow plugs are briefly operated in this phase with push voltage, which is above the glow-plug nominal voltage. The activating voltage is then reduced to the glow-plug nominal voltage during standby heating.

During start heating, the activation voltage is increased again in order to compensate for the cooling of the glow plug by the cold intake air. This is also possible in the post- and intermediate-heating phases. The required voltage is taken from a number of program maps that are adapted to each engine. The program maps contain among others the parameters engine speed, injected fuel quantity, time after starter release, and coolant temperature.

Map-controlled activation reliably prevents thermal overloading of the glow plugs in all engine operating states. The heating function implemented in the engine ECU contains an overheating-protection facility in the event of repeat heating.

These low-voltage preheating systems facilitate a rapid start similar to a gasoline engine down to approximately –30 °C.

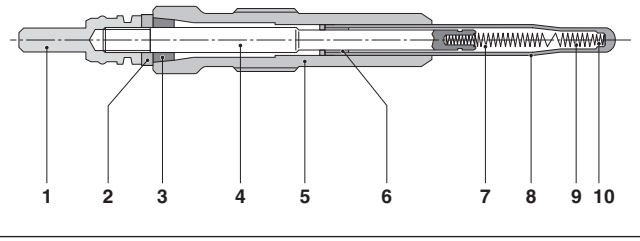
Low-voltage metal glow plug

Design and characteristics

The main component in the sheathed-element glow plug is the tubular heating element (Figure 3). The tubular heating element consists of a hot-gas and corrosion-resistant element sheath which encloses a filament surrounded by compressed magnesium-oxide powder. This filament is made up of two resistors con-

Figure 3: Glow plug

1 Plug connector,
2 Insulator shim,
3 Double seal,
4 Terminal stud,
5 Housing,
6 Element seal,
7 Control filament,
8 Glow tube,
9 Magnesium-oxide powder (filling powder),
10 Heating filament.



nected in series – the heating filament located in the tip of the sheath and the control filament.

Whereas the heating filament has an electrical resistance that is independent of temperature, the control filament has a positive temperature coefficient (PTC). Its resistance increases as the temperature rises.

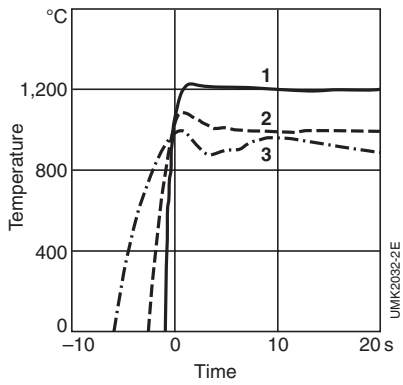
The heating filament is welded into the cap of the element sheath for grounding. The control filament is contacted at the terminal stud, which establishes the connection to the vehicle electrical system.

The glow element is tapered at the front in order to accommodate the heating filament closer to the element sheath. This allows preheating rates of up to 1,000 °C within 3 s with the push mode (Figure 4). The maximum heating temperature is in excess of 1,000 °C. The temperature during start-standby heating and in post-heating mode is approx. 980 °C.

Glow plugs with 11 V nominal voltage are practically no longer used in passenger cars due to their longer heat-up time.

Figure 4: Comparison of preheating curves, from t = 0 s with a flow velocity of 11 m/s

1 Low-voltage ceramic glow plug (e.g., 7 V nominal voltage),
2 Low-voltage metal glow plug (e.g., 5 V nominal voltage),
3 Metal glow plug (11 V nominal voltage).



Operation
When voltage is applied to the glow plug, most of the electrical energy of the heating filament is initially converted into heat. The temperature at the tip of the glow plug increases sharply. The temperature of the control filament – and with it also the impedance – increase with a time delay. The current draw and thus the total heating output of the glow plug decrease and the temperature approaches a steady-state condition. By controlling the operating voltage the temperature is kept below the critical level for the glow plug. Consequently, it can remain in operation for several minutes after the engine has started. This post-heating function results in a more effective engine cold-idle phase with substantially lower noise and emission output.

Reduced emissions in diesel engines with a low compression ratio

By lowering the compression ratio in modern diesel engines from $\varepsilon = 18$ to $\varepsilon = 16$, it is possible to reduce the NO_x and soot emissions while simultaneously increasing the specific power. However, the cold-starting and cold-idling performance is more difficult to control in these engines due to the lower compression. Temperatures at the glow plug of up to 1,250 °C are required in order to achieve minimum exhaust-gas opacity values and smooth running during cold starting and cold idling of these engines. Optimum combustion can only be achieved during the cold-running phase by post-heating lasting several minutes. These requirements can barely be satisfied with metal glow plugs.

Low-voltage ceramic glow plugs

Low-voltage ceramic glow plugs have glow elements made of a ceramic material that is highly resistant to temperatures. Ceramic heating-conductor tracks are embedded into the ceramic material. Their function is comparable to the filaments of metal glow plugs.

Due to their very high resistance to oxidation and thermal shocks, ceramic glow plugs enable an immediate start, as well as minute-long post-heating and inter-

mediate heating at up to 1,250 °C. They are designed for a nominal voltage of 7 V.

Glow control unit

The glow control unit sets the temperature at the glow plug stipulated by the engine management for the respective operating state. The supply voltage is pulse-width-modulated via a transistor for this purpose.

The glow control unit furthermore assumes functions like ground-offset compensation and protection against polarity reversal.

Applications

Low-voltage preheating systems require an application of the system matched to the respective engine in order to set the best temperature in each operating state and simultaneously protect the glow plug against being damaged by overvoltage. The respective activation voltage is stored in program maps.

Regulated preheating system

The regulated preheating system was developed to minimize the application outlay. Ceramic glow plugs exhibit within certain limits a linear behavior between their temperature and their resistance. This behavior is used by an algorithm to calculate in the ECU the current temperature actually applied at the glow element and to automatically adjust the desired temperature accordingly.

Flame start systems for commercial-vehicle diesel engines

Requirements

Commercial-vehicle diesel engines can start down to temperatures of around –20 °C with start assistance. The start durations are dependent on the compression ratio ε and the fuel grade used (cetane number CN, Figure 5). To prevent misfiring and emissions of unburnt hydrocarbons (HC), it is however possible, depending on the compression ratio and the fuel grade used, to utilize already from a coolant temperature of 10 °C air preheating as start assistance (flame start system or grid heater).

Air preheating provides for safe, quick starting down to low outside temperatures. The battery and the starter are protected and fuel consumption is reduced in the starting and warm-up phases. At higher temperatures air preheating would not be necessary for safe starting, but does prevent the formation of white smoke (HC emission, Figure 6).

Figure 5: Starting times without start assistance as a function of compression ratio ε and cetane number CN

1 Starting-time curve for ε = 16.5 and CN = 41,
2 Starting-time curve for ε = 16.5 and CN = 45,
3 Starting-time curve for ε = 16.5 and CN = 51,
4 Starting-time curve for ε = 17.0 and CN = 51,
5 Starting-time curve for ε = 17.75 and CN = 51.

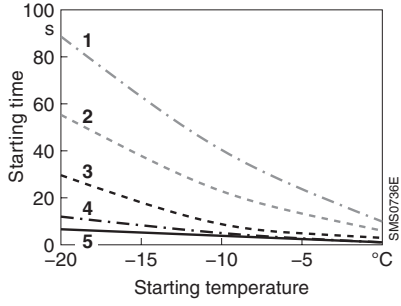
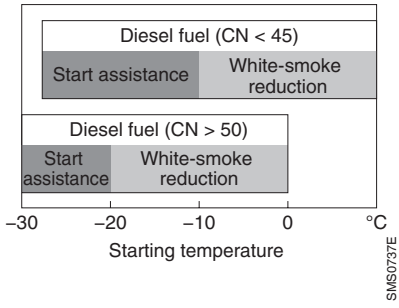


Figure 6: Use of start assistance for starting and to reduce HC emissions



Design

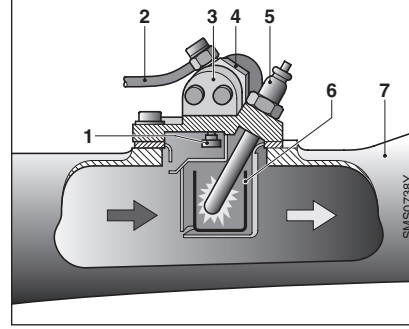
There are two different types of flame start system: the flame start system with burner chamber and the flame glow plug.

Flame start system with burner chamber

The flame start system with burner chamber (Figure 7) consists of a nozzle holder on which a solenoid valve, a nozzle and a rod-type glow plug are mounted. The nozzle holder is attached together with gaskets and the burner chamber by three screws to the charge-air housing.

Figure 7: Flame start system with burner chamber

1 Nozzle, 2 Fuel line,
3 Nozzle holder, 4 Solenoid valve,
5 Rod-type glow plug,
6 Burner chamber,
7 Charge-air housing.

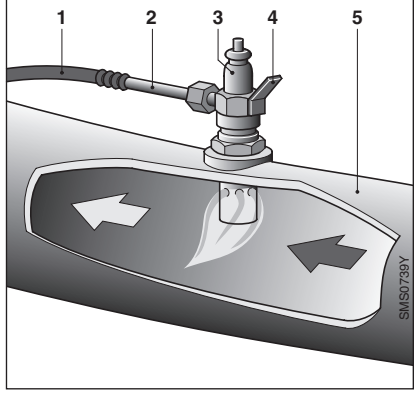


The fuel is drawn after the fuel-system fuel filter and routed to the solenoid valve via a fuel line. The solenoid valve controls the fuel flow while the flame start system is in operation. The glow plug is supplied with electrical power by the flame-start control unit.

Flame glow plug
In contrast to the flame start system with burner chamber, the solenoid valve in the case of the flame glow plug is not mounted directly on the flame start system, but instead is connected by a fuel line to the flame glow plug. The flame glow plug also has a much more compact design (Figure 8).

The glow plug integrated in the flame glow plug is enclosed in a vaporizer tube in the upper area and extends up to the protective tube at the end of the flame glow plug (Figure 9). The fuel is drawn after the fuel filter, routed through the separately mounted solenoid valve, and supplied to the flame glow plug via the fuel line. The fuel enters the flame glow plug via the metering device – to control the fuel quantity. The flame-start control unit assumes the job of controlling the flame start system (current supply to the glow plug and switching on the solenoid valve).

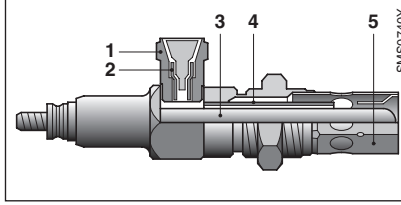
Figure 8: Installed flame glow plug
1 Fuel line,
2 Throttle line,
3 Flame glow plug,
4 Ground connection,
5 Charge-air tube.



Use of flame start systems
The flame start system with burner chamber is used primarily in commercial-vehicle engines with a displacement $V_H > 10\,l$, while the flame glow plug is used in engines with a displacement V_H of 4 to 10 l. The flame start system is installed in the engine charge-air tube in such a way that the heated intake air is uniformly distributed to all the cylinders.

Operating principle
The flame start system is activated below the switch-on temperature defined in the control unit. When the ignition key is turned to the Drive position, the "Flame start system" indicator lamp lights up and the glow plug is energized and preheated. The indicator lamp goes out after a preheating time dependent on the vehicle system voltage (approx. 25 s). The flame start system is operational and the engine can be started. At the outset of the starting operation the solenoid valve opens and the fuel flows to the hot glow plug, where it ignites. The flame that is now produced heats the intake air flowing past. The flame start system is operated after the engine start in the post-flaming time until a stable burn is guaranteed and shuts down according to the parameters defined in the control unit. The parameters for post-flaming are dependent on the coolant temperature.

Figure 9: Flame glow plug (Beru system)
1 Fuel-supply connection,
2 Metering device,
3 Glow plug,
4 Vaporizer tube,
5 Protective tube.



The flame start system is also shut down if the engine is started before the indicator lamp goes out or is not started within 30 seconds of the indicator lamp going out. In this case the driver can put the flame start system into operation again by turning the ignition key back and then turning into to the drive position again.

Control of the glow plug

Glow plug for the flame start system with burner chamber

The glow plug is preheated depending on the vehicle system voltage until a temperature of approximately 1,050 °C is reached. This is followed by the beginning of the period of start readiness (duration 30 s), during which the glow plug is cyclically energized. The cycle ratio is set depending on the current vehicle system voltage in such a way that the glow-plug temperature remains between 1,020 °C and 1,080 °C. With the glow plug in this temperature range the fuel is safely ignited and the glow plug itself is not exposed to thermal overload.

As soon as the engine is started, the solenoid valve opens, the fuel is directed through the nozzle and sprayed onto the spark-plug grooves, where it ignites. In this phase the cycle ratio of the glow plug is increased by the control unit so that the glow plug does not cool in response to the heat dissipated from the vaporizing fuel. This higher cycle ratio is maintained after starting for a further period of around 30 s until all the components are heated and the flame is safely burning. The glow plug is now additionally heated by the flame. The control unit decreases the cycle ratio again to protect against overheating. After post-flaming the glow plug is deactivated and monitored until it has cooled down entirely. Start and cold-running support by the flame start system is concluded.

Flame glow plug

This glow plug is preheated in exactly the same way before starting, depending on the applied vehicle power supply. The preheating times are slightly longer than for the glow plug with burner chamber, since the vaporizer tube is additionally heated during preheating (see Figure 9). When the engine is started, the solenoid valve is opened, fuel flows via the fuel-supply connection and the metering device into the flame glow plug, strikes the hot glow plug, vaporizes between the glow plug and the vaporizer tube, and mixes with the air at the evaporator-tube outlet. The air/fuel mixture ignites at the hot glow plug. The flame produced emerges from the protective tube and heats the intake air flowing past.

After preheating the flame glow plug is likewise cyclically energized. The cycle ratio is decreased in the start-readiness period (30 s) and increased again when the engine is started. In this phase the cycle ratio for the flame glow plug is higher than for the flame start system with burner chamber, since the fuel is vaporized via the supplied electrical energy and not via the burning flame.

Cooling-down phase

The glow plug of the flame start system is monitored in the cooling-down phase. If preheating is performed again in this phase, the control unit calculates the preheating time from a program map depending on the cooling-down time and the voltage in such a way that the glow plug reached the temperature range described above.

Because the glow plugs have heating and control filaments (see Preheating systems for passenger-car diesel engines), the optimum preheating time is set as a function of the ambient conditions.

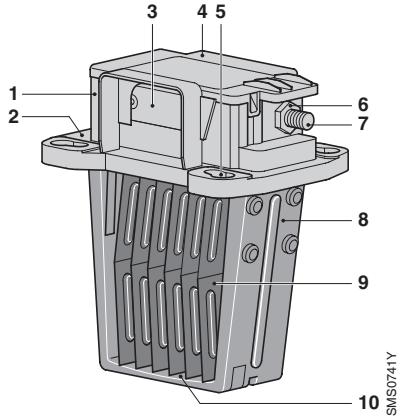
Grid heater for commercial-vehicle diesel engines

The grid heater is used as a further cold-start aid, particularly for engines which are started under aggravated starting conditions such as constant load when starting (e.g. for hydraulic output) or at altitudes above 2,000 m (e.g. snow groomers).

Design
The grid heater (Figure 10) consists of an electric heater element which is installed in the charge-air tube and powered by a relay. A separate ECU or the engine ECU switches on the grid heater on the basis of the stored parameters.

There are different versions which are used, depending on the engine displacement V_H and the vehicle system voltage (12 V or 24 V). Grid heaters with a power input of 1.8 kW are used for engines with a displacement V_H of 4 to 10 l, while grid heaters with a power input of 2.7 kW are used for engines with $V_H > 10\,l$.

Figure 10: Grid heater
1 Housing,
2 Mounting flange,
3 Cover plate,
4 Cover,
5 Mounting bore,
6 Lock nut,
7 Connecting-thread bolt,
8 Retaining frame,
9 Heating filament,
10 Insulating plate.



Operating principle
Depending on the coolant and charge-air temperatures, the grid heater is preheated for periods varying in duration (preheating time t_v = 2 to 28 s) and remains active in the post-heating phase until stable engine running is ensured. The switch-on temperature is 5 °C for engines with V_H = 4 to 10 l and – 4° for engines with $V_H > 10\,l$.

Once the preheating time has elapsed, the indicator lamp goes out and the engine can be started. On starting the grid heater is switched off for approximately 2 s so as to make full battery capacity available to the starter. For start and idle support the grid heater is then switched back on and remains in operation until it is switched off after the applied post-heating time.

The grid heater remains in start readiness for only around 5 s (as opposed to 30 s for the flame start system) in order to conserve battery capacity.

Advantages and disadvantages of the grid heater
The grid heater offers the following advantages over the flame start system:
– The intake air is heated before starting
– There are shorter starting times thanks to good ignition conditions from the first revolution
– No burning of oxygen is required to heat the intake air

On the down side, the grid heater requires greater battery capacity.

Exhaust-gas treatment

Catalytic oxidation

The oldest catalytic method of exhaust-gas treatment is aimed at the oxidation of imperfectly combusted fuels. Carbon monoxide (CO) and imperfectly combusted hydrocarbons ($C_xH_yO_z$) – which also include soot and hydrogen (H_2) – are converted in this process with the oxidizing agents present in the exhaust gas, i.e. oxygen (O_2) in particular. These reactions correspond to the stoichiometries of the equations (A) through (D):

(A) $C_x + x\,O_2 \rightarrow x\,CO_2$,
(B) $2\,CO + O_2 \rightarrow 2\,CO_2$,
(C) $2\,H_2 + O_2 \rightarrow 2\,H_2O$,
(D) $4\,C_xH_yO_z + (4x+y-2z)\,O_2 \rightarrow 4x\,CO_2 + 2y\,H_2O$.

As a rule noble metals, their mixtures or their alloys in the form of nanoparticles serve as the catalytically active component. Platinum and palladium are frequently used.

A prerequisite for the most complete sequence of oxidation reactions as possible is an at least stoichiometric quantity of oxygen in the exhaust-gas flow. A catalytic converter can however passivate with high oxygen excess and low exhaust-gas temperatures. The elimination of imperfectly combusted fuels from the exhaust gas works particularly effectively when the catalytic converters and the engine application are matched to each other. The following conditions should be in place for this purpose:

– The engine exhibits low untreated emission of imperfectly combusted fuel, particularly in the cold-start phase.
– The engine application provides the catalytic converter with a balanced air/fuel ratio (lambda, λ).
– The oxidation catalytic converter is heated as quickly as possible by means of heating measures.
– The catalytic converter has a quantity of catalytically active and storing components matched to the emissions and exhaust-gas temperatures.
– The type and quantity of all the components are matched in such a way that the catalytic converter resists to a sufficient extent the loads of the vehicle application (thermal load, sulfur and other contaminations) over the aimed-for service life.

In addition, with an appropriate choice of components in such oxidation catalytic converters, nitrogen monoxide can be oxidized in accordance with reaction equation (E). As well as the importance of NO_2 to the SCR reaction (see Selective catalytic reduction) the emitted particulate matter can be reduced in accordance with the stoichiometry (F) with the created NO_2. Essentially, the hydrocarbons present in the particulates are oxidized using NO_2 in accordance with stoichiometry (D):

(E) $2\,NO + O_2 \rightarrow 2\,NO_2$,
(F) $C_x + 2x\,NO_2 \rightarrow x\,CO_2 + 2x\,NO$.

Without further exhaust-gas treatment measures it is not possible for a large number of vehicles to comply with strict particulate-emission limits (see for example EC Regulation 582/2011 [1]).

Filtration of particulates

To comply with the emission-control standard Euro 6c, a passenger car, irrespective of whether it is powered by a direct-injection gasoline engine or by a diesel engine, may not exceed the limit value for the particulate matter (PM) of 4.5 mg/km and the limit value for the particle number (PN) of $6 \cdot 10^{11}$ particles/km.

Particulate filter
Many vehicles need a particulate filter to achieve these limit values (Figure 1). These filters are for the most part ceramic honeycomb structures in which all the channels are alternately sealed at on end. Unlike other catalytic-converter honeycombs, it is therefore not possible to look through particulate filters. The exhaust gas must as it passes the honeycomb channels flow through the porous wall material (Figure 2). The pore structure of the ceramics is designed to accommodate the size of the particles. The most common particulate filters consist of silicon carbide (SiC), aluminum titanate (Al_2TiO_5), or cordierite ($Mg_2Al_3[AlSiO_5O_{18}]$). There are also special forms made of sintered metal.

Some of the residues of engine oil, fuel, combustion air, and engine attrition that collect in the particulate filter cannot be removed by combustion and, if suitably sized, remain as ash in the filter. On the other hand, soot and the organic compounds that adhere to it are combustible.

NO_2 and O_2 work as oxidizing agents. Oxidation with NO_2 works at exhaust-gas temperatures from around 200 to 250 °C. From around 450 °C NO prevails in the thermodynamic equilibrium of the nitrous fumes. Then there is little NO_2 in the exhaust gas.

Catalytically oxidizing coatings of particulate filters do not differ in their type and function fundamentally from other oxidation catalytic converters. They serve to set a favorable NO-NO_2 ratio, oxidize imperfectly combusted fuel, and catalyze the CO that is produced during soot burn-off.

Regeneration of the particulate filter
Soot is oxidized by oxygen only at temperatures above approximately 550 to 600 °C. Such temperatures exist in many diesel engines only with very high power outputs and hardly ever arise during the normal operation of some vehicles. Measures must therefore be taken to raise the exhaust-gas temperature. Combustion control of the engine, for example, is changed for this purpose.

Thermal measures
Important engine measures to increase exhaust-gas temperature are advanced burnoff or additive secondary injection, retarded main injection, retarded secondary injection, and intake-air throttling.

Figure 1: Diesel particulate filter
1 Housing,
2 Extruded honeycomb ceramic,
3 Ceramic plug.

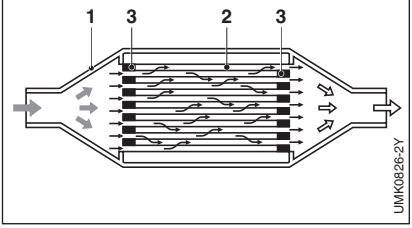
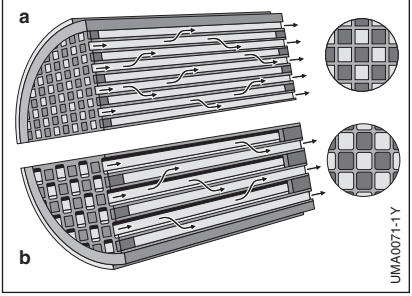


Figure 2: Designs of ceramic particulate filter
a) Square channel cross-section,
b) Octosquare design.



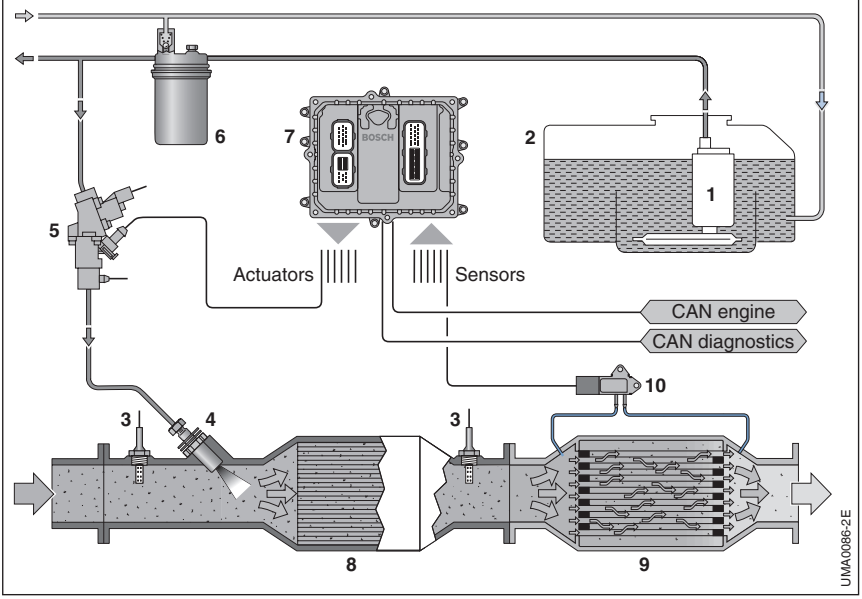
Depending on the engine operating point, one or several of these measures are triggered.

HCI system
As an alternative to injecting additional fuel into the engine cylinders, fuel can also be metered directly into the exhaust-gas system. The HCI system (Hydrocarbon Injection) serves this purpose. Here, the diesel fuel is for the most part injected ahead of an oxidizing catalytic converter and vaporized (Figure 3). The fuel vapors oxidize in the catalytic converter, resulting in an increase in the temperature of the exhaust-gas flow. The control algorithms of the HCI system are associated with soot-mass determination.

Additive system
In addition to the thermal measures to raise the exhaust-gas temperature, the ignition temperature of soot can also be reduced. This is done by catalytically active components which are finely distributed into the soot. These components are supplied as a fuel additive (Figure 4). Organometallic compounds are used, for example ferrocene. The finely distributed metal oxides formed from them reduce the ignition temperature of soot by up to 150 K. One drawback is the ash that is created from this fuel additive.

Saturation detection
Important factors in the design of particulate filters and their regeneration strategy are the amount of ash to be expected, the maximum soot load until regeneration is initiated, and the permissible exhaust-gas back pressure.

Figure 3: HCI system (Hydrocarbon Injection)
1 Fuel pump, 2 Fuel tank, 3 Temperature sensor, 4 HC metering module, 5 HC metering unit, 6 Fuel filter, 7 Engine ECU, 8 Diesel oxidation catalytic converter, 9 Diesel particulate filter, 10 Differential-pressure sensor.



Often two processes are used in parallel for saturation detection. The particulate filter's flow resistance is calculated from the pressure drop of the exhaust-gas flow detected with a differential-pressure sensor over the particulate filter. It can be related to the amount of soot stored in the particulate filter. In addition, a model is used to calculate the soot mass stored in the particulate filter. To this end, the soot-mass flow of the engine is integrated, where different corrections can be taken into consideration, e.g. for the soot oxidation of the particulates by NO_2.

Catalytic reduction of nitrous fumes

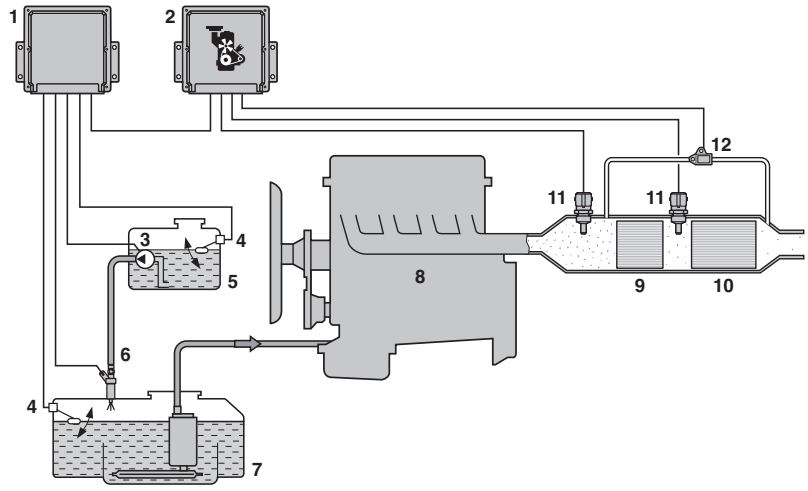
Requirements

Catalytic methods for reducing the NO_x emissions to comply with statutory requirements work particularly when vehicle design, engine application, and the components of the exhaust-gas treatment system are matched to each other. The following points must be taken into consideration here:

- The vehicle offers enough spacer to accommodate the necessary exhaust-gas treatment system. Special requirements with regard to the positioning of individual components in the vehicle must be observed (e.g. positioning near the engine).
- Exhaust-gas temperature, exhaust-gas mass flow, and exhaust-gas composition (NO_x, λ etc.) must be selectively and quickly regulated.
- The engine or a component of the exhaust system makes it possible to provide the amount of reducing agent

Figure 4: Particulate filter with additive system

1 Additive control unit, 2 Engine control unit, 3 Additive pump, 4 Fill level sensor, 5 Additive tank, 6 Additive injection nozzle, 7 Fuel tank, 8 Diesel engine, 9 Oxidation catalytic converter, 10 Particulate filter, 11 Temperature sensor, 12 Differential pressure sensor.



needed to reduce the untreated NO_x emissions.

The choice of a method of reducing NO_x suitable for a concrete vehicle application is therefore always a consideration of the complete system. The two NO_x storage catalytic converters and NO_x reduction catalytic converters described below can also be used together, for example to achieve high conversion rates.

NO_x storage catalytic converter

The reduction of NO_x from the exhaust gas of diesel engines works under suitable conditions with an NO_x storage catalytic converter (NSC, NO_x Storage Catalyst; alternatively called LNT, Lean NO_x Trap) very similar to the three-way catalytic converter (TWC, Three-Way Catalyst) (Figure 5).

Engine-based λ control constitutes an alternation of oxygen-richer and low-oxygen exhaust-gas sequences for this purpose. The oxygen-enriched phase often lasts several minutes. An accordingly large amount of storage components is required to bind the NO_x created during this period. The NO_x previously bound to the storage components is reduced in the subsequent fuel-enriched sequence. This sequence, known as NSC regeneration, usually takes a few seconds.

The storage components desorb NO_x at high temperatures also without the exhaust gas being enriched.

The NSC method works in particularly at exhaust-gas temperatures of between around 180 °C and 400 °C. The NO_x conversion of the NSC method deteriorates at significantly higher temperatures.

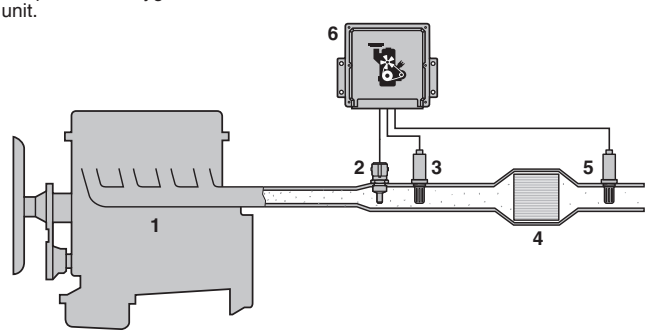
Contamination by sulfur reduces the efficiency of the noble metals and the NO_x storage components. A storage catalytic converter must therefore be regularly desulfurated. High temperatures are required for this purpose. When it comes to the design of NO_x storage catalytic converters, particular attention must be given to their long-term behavior depending on the relevant engine application with the details of desulfuration.

Selective catalytic reduction

SCR method

The SCR method (Selective Catalytic Reduction) is based on the principle of adding a reducing agent to the exhaust gas (Figure 6). Diesel fuel is basically also suitable here. However, it has hitherto not been possible to develop a suitable SCR method based on diesel fuel as the reducing agent that provides for compliance with strict emission limits.

Figure 5: Schematic diagram of an exhaust-gas system with NO_x storage catalytic converter
1 Diesel engine, 2 Temperature sensor,
3 Wide-band λ sensor (LSU),
4 NO_x storage catalytic converter,
5 NO_x sensor or two-position λ oxygen sensor,
6 Engine control unit.



Gaseous ammonia (NH_3) is used in stationary combustion systems, e.g. in power stations, in the SCR method. NH_3 gas is avoided in motor vehicles for safety reasons. Instead, in motor vehicles for the most part an aqueous solution of urea ($(NH_2)_2CO$) is metered into the exhaust-system branch in the direction of flow ahead of the SCR catalytic converter (Figure 6). Such aqueous solutions are called for example AdBlue, DEF (Diesel Exhaust Fluid), or AUS 32 (Aqueous Urea Solution) (see urea/water solution).

Aside from special forms, SCR catalytic converters frequently consist of catalytically active coatings on ceramic honeycomb structures. Special ceramic particulate filters are also suitable for such coatings. Vanadium oxides and iron and copper zeolites are suitable for use as the catalytically active material.

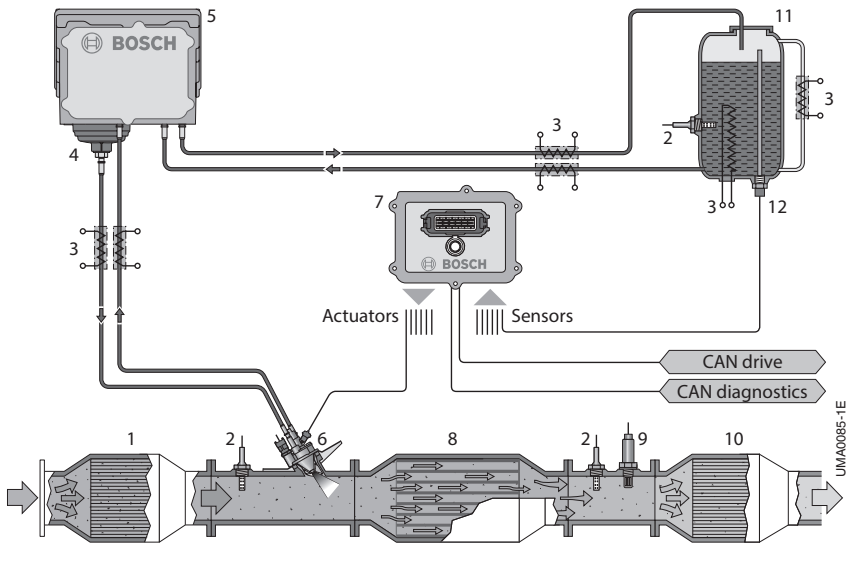
Reaction equations

Ammonia reacts on the SCR catalytic converter with NO_x according to the following equations:

(G) $4\ NO + 4\ NH_3 + O_2 \rightarrow 4\ N_2 + 6\ H_2O$,
(H) $NO + NO_2 + 2\,NH_3 + H_2O + O_2 \rightarrow 2\,NH_4NO_3$,
(I) $NH_4NO_3 \rightarrow N_2O + 2\ H_2O$,
(J) $2\ NH_4NO_3 \rightarrow 2\ N_2 + 4\ H_2O + O_2$,
(K) $NO + NO_2 + 2\ NH_3 \rightarrow 2\ N_2 + 3\ H_2O$,
(L) $6\ NO_2 + 8\ NH_3 \rightarrow 7\ N_2 + 12\ H_2O$.

The reaction according to equation (G) is comparatively slow. It predominates if no or only little NO_2 is available to react with NH_3. If the energy of activation is not achieved to reduce the nitrous fumes, ammonium nitrate (NH_4NO_3) is created according to equation (H) in a secondary reaction. NH_4NO_3 decomposes at temperatures from approximately 170 °C according to equation (I). If it is heated very quickly, it decomposes ex-

Figure 6: Exhaust-gas system with catalytic reduction of nitrogen oxides (SCR)
1 Diesel oxidation catalytic converter, 2 Temperature sensor, 3 Heater, 4 Filter, 5 Delivery module, 6 Metering module, 7 Metering control unit, 8 SCR catalytic converter, 9 NO_x sensor, 10 Slip catalytic converter, 11 Urea-solution tank, 12 Urea-solution fill-level sensor.



plosively according to equation (J). If the exhaust-gas flow contains equal proportions of NO and NO_2, the faster reaction predominates according to equation (K) if its energy of activation is reached. Where there is a higher proportion of NO_2 in the exhaust gas, the reaction can according to equation (L) contribute significantly to the overall conversion.

Choice of SCR catalytic converter
The engine emissions and the NO_x conversion needed to achieve the emission limit and the exhaust-gas temperature in the relevant operating states are decisive factors in the choice of a suitable SCR catalytic converter. However, the sulfur content of the fuel used is also important. When it comes to the design of the SCR catalytic converter, attention must also be given in particular to its long-term stability depending on the concrete vehicle application, e.g. also with the details of desulfuration.

Metering of the reducing agent
SCR catalytic converters do have a certain NH_3 storage capacity, but essentially a quantity of urea/water solution equivalent to the instantaneous NO_x mass flow must be metered as precisely as possible and dispensed into the exhaust-system branch. If more reducing agent is dispensed than is converted in reduction with NO_x, it may result in unwanted NH_3 slip. This excess NH_3 can be oxidized by an additional oxidation catalytic converter at the end of the SCR catalytic converter. In the most favorable case, nitrogen and water are created at this NH_3 slip catalytic converter. Nitrogen oxides are created in secondary reactions in unfavorable circumstances.

Before entering each catalytic converter, the exhaust-gas mass flow must be distributed as uniformly as possible to the individual honeycomb channels. Uniform distribution of the fluids requires particular attention in the case of two-substance mixtures, i.e. if a second substance is added to the exhaust-gas flow, as is the case with the SCR method.

To facilitate the SCR reactions, ammonia must first be created from the aqueous urea solution. This occurs in a chronological sequence of different steps:
– determination of the metering quantity,
– delivery of the solution,
– dispensing of the metered volume,
– atomization and vaporization of the solution,
– decomposition of urea, and
– distribution of ammonia through the catalytic-converter cross-section.

The metering quantity is determined in a separate ECU of the exhaust-gas treatment systems (optionally integrated in the engine control unit) on the basis of stored data and data determined by exhaust-gas sensors (temperature sensor, NO_x sensor, NH_3 sensor, etc.). In the process the current NH_3 storage load of the catalytic converter is also calculated and taken into account as a rule. The data required for this calculation are determined in advance for the catalytic converter used in each case and stored as a program map in the ECU. To meter the required quantity as precisely as possible, it is necessary to bear in mind that ISO 22241-1 [2] defines the urea content of these solutions at 31.8 to 33.2 % by mass. Significant deviations in the urea-solution concentration can also arise as a result of their aging in the vehicle or incorrect storage. Quality sensors are therefore increasingly being used.

Delivery and metering modules
A delivery module as a rule consists of a liquid pump, e.g. a diaphragm pump. It provides the solution from the urea tank for metering. Since commercially available urea solutions freeze below –11 °C, all the components that are in contact with the solution must be able to withstand ice pressure. Some components have heating elements to melt the ice where necessary. The metered volume is dispensed via metering modules. These are primarily liquid nozzles, or more rarely air/fluid two-substance nozzles. The metering modules in contact with the hot exhaust

gas must be cooled to prevent the solution from vaporizing prematurely.

Vaporization of the solution is preceded by the distribution of the droplets sprayed into the exhaust-gas flow. The spray characteristic, i.e. the droplet size and its impulse, must be closely matched with temperature and flow conditions in the affected part of the exhaust-system branch. As a rule some of the droplets vaporize on hot surfaces of the exhaust system, e.g. on pipe walls or on gas mixers.

Release of ammonia
Vaporization of the droplets is followed by the release of ammonia due to the decomposition of urea. In the first thermolytic decomposition stage of urea, initially isocyanic acid (HNCO) is created as well as NH_3 under the conditions prevailing in the exhaust-system branch according to reaction equation (M):

(M) $(NH_2)_2CO \rightarrow NH_3 + HNCO$.

This decomposition of urea occurs above its melting temperature from approximately 133 °C. Under suitable conditions the isocyanic acid created in the first stage is subjected in a second stage to hydrolysis with the creation of ammonia and carbon dioxide according to equation (N):

(N) $HNCO + H_2O \rightarrow NH_3 + CO_2$.

This comparatively slow hydrolysis reaction is significant in the gaseous phase merely in the case of a long retention time of very small droplets, which vaporize already in the gaseous phase at accordingly high temperatures. Hydrolysis of isocyanic acid therefore occurs for the most part on surfaces. As well as the most uniform distribution as possible of the created ammonia into the exhaust-gas flow and its distribution as perfectly as possible through the catalytic-converter cross-section, it must also be borne in mind when designing that part of the exhaust-system branch into which the agent is metered that solid by-products may be created when urea decomposes. Surface temperatures of well over 300 °C are needed to eliminate their deposits.

On-Board Diagnostics (OBD)

OBD describes self-diagnostic systems for testing essential vehicle functions, including exhaust-gas treatment. A multitude of sensors and comprehensive software supply most of the information for these diagnoses. OBD has been subject to continuous further development over the past (more than) 20 years. Today OBD systems influence the vehicle's service-inspection program and the driver and the workshop are informed of malfunctions and recommended actions. The OBD requirements applicable in each case must be taken into account in the design of the exhaust-gas treatment system.

References
[1] Commission Regulation (EU) No. 582/2011 of 25 May 2011 implementing and amending Regulation (EC) No. 595/2009 of the European Parliament and of the Council with respect to emissions from commercial vehicles (Euro VI) and amending Annexes I and III to Directive 2007/46/EC of the European Parliament and of the Council (Text with EEA relevance).
[2] ISO 22241-1: Diesel engines – NO_x reduction agent AUS 32 – Part 1: Quality requirements.

Electrification of the drive

Features

Definition Electromobility
Electromobility refers generally to that part of mobility for which electrical energy or power is used. It includes for example the operation of electrified trains and riding with electrically assisted bicycles. With regard to motor vehicles the following vehicles are assigned to electromobility in this Handbook:
– battery electric vehicles,
– vehicles with fuel cell,
– externally rechargeable hybrid vehicles, and
– hybrid vehicles which can cover significant distances by electric-only means.

This corresponds to a large extent to the definition from the Gabler Dictionary of Economics [1]). With regard to other definitions, only vehicles with an electric motor (electrical machine, e-machine) which draw their energy primarily from the mains-supply network may be assigned to electromobility. Consequently, fuel-cell vehicles which are run on hydrogen as their energy source are not assigned to electromobility.

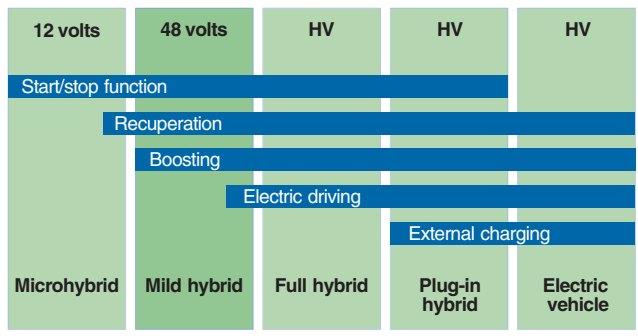
Hybrid vehicles which can only cover short distances by electric means whose electrical machine therefore only has the task of supporting the internal-combustion engine are on no account assigned to electromobility.

Electric vehicles and hybrid vehicles in their different design variations have an electrical machine which acts in some way on the drivetrain. They are therefore referred to as electrified drives.

Development of electric drives in motor vehicles
History of electric drives
Electric drives in motor vehicles are not an invention of the last few years. The first carriage recognized as an electric vehicle was a tricycle by Gustave Trouvé in 1881. The first German vehicle with an electric drive was the Flocken electric car of 1888. In this period numerous manufacturers around the world developed properly functioning electrically powered vehicles. They were in competition with vehicles with internal-combustion engines, which were also in the initial stages of development.

The golden age of electric cars was from 1896 to 1912. The range of electric vehicles on a single battery charge was around 100 km. To increase the range, vehicles were developed in which a gasoline engine via an alternator charged the battery, which powered the electric motor – i.e., hybrid electric vehicles.
After 1912 the market share of electric vehicles dropped off dramatically. The

Figure 1: Variants of electrified drives for motor vehicles
HV High voltage.



reasons for this were the much greater range of vehicles with internal-combustion engines, the cheap fuel gasoline, the now comfortable process of starting the internal-combustion engine with the starter, but also the problem with the traction batteries.

Rediscovery of the electrified drive
The electrification of the drive started to gather momentum only at the end of the 20th century with the demand for vehicles with lower exhaust-gas emissions and the demand for reduced fuel consumption. The high fuel consumption of internal-combustion engines results from the poor degree of efficiency of around 40 %. Electrical machines are at an advantage with an efficiency in excess of 90 %. However, the high-volume application of electric drives powered by one battery with a suitably large battery capacity initially failed due to the embryonic battery technology and the high costs. Because there was minimal take-up on electric vehicles owing to their limited range on one battery charge, vehicles with hybrid drives – i.e., a combination of electric drive and internal-combustion engine – were offered as a technology to bridge the gap.

Advances in battery development resulted in a significant increase in the range. In the meantime a series of electric vehicles is available on the market and a sharply growing market share of electric vehicles is anticipated in the future.

Variants of electrified drives for motor vehicles
The following variants of electrified drives can be found in motor vehicles (Figure 1):
– The mild hybrid with 48V battery assists the combustion engine (boosting) and offers the possibility of maneuvering the vehicle on electric power.
– The full hybrid with traction battery with a battery voltage of typically 400 V (HV, High Voltage) offers the possibility of driving limited distances by electric-only means with zero emissions locally.
– The external rechargeable hybrid can be recharged for example at the power socket in the garage (e.g., plug-in hybrid). The journey can thus be started with a fully charged battery; longer distances can by covered on electric-only power with a larger battery compared with the full hybrid.
– The electric drive with a traction battery (battery voltage typically 400 V or 800 V for powerful vehicles and commercial vehicles) propels the vehicle on electric-only power.
– A vehicle known as a “range extender” has an electric drive and also for example a small internal-combustion engine. It is designed as an electric drive for electric-only driving, with the corresponding performance of the electrical machine and traction battery. If the state of charge of the traction battery is no longer sufficient for propulsion, it is recharged by the combustion engine via a second electrical machine so that the vehicle can continue to be driven at reduced speed. In this operating mode this drive becomes a serial, externally rechargeable hybrid. Other concepts provide for the combustion engine to be mechanically connected directly to the drive (parallel hybrid). The combustion engine acts as a range extender.
– The electric drive with fuel cell produces the energy for propulsion continually from hydrogen and oxygen. A small back-up battery is used to intermediately store the generated energy. With a large battery this can be used for the drive; a small fuel cell charges the battery as required and thus acts as a range extender.

The electrical machine of the microhybrid set out in Figure 1 is not designed power-wise to make a contribution to the drive; this variant therefore is not a genuine hybrid.

Components of the electrified drive

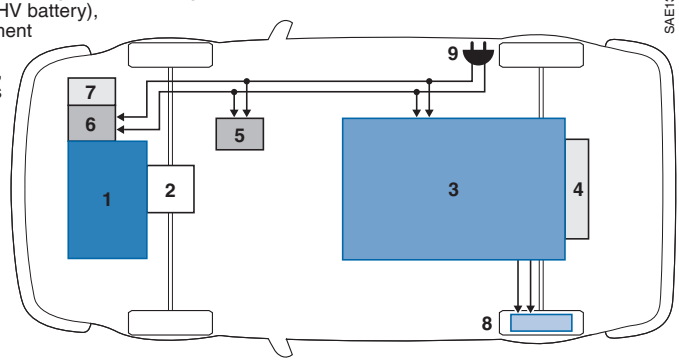
Electric drive
An electric drive consists of the following components (Figure 2):
– electrical machine (e-machine),
– traction battery with battery-management system,
– power electronics for activating the electrical machine (inverter),
– electronic control (VCU, vehicle control unit),
– DC/DC converter,
– charging unit with charging socket, and
– transmission (single-stage reduction-gear unit or multi-speed transmission for higher power demand).

Consequently, the drivetrain of an electric vehicle consists of much fewer components compared with a conventional vehicle with combustion engine.

Interaction of the components
The traction battery (see Batteries for electric and hybrid vehicles) is the energy source for the battery-electric-powered electric vehicle; the energy is stored by charging the battery. In fuel-cell vehicles the current is generated continually (see Fuel cell). Both energy sources supply direct current. To operate the electrical machine the direct current must be converted into a three-phase alternating current (see Electrical machines for drives). This task is performed by an inverter, which activates the electrical machine in line with demand by targeted switching of the direct current (see Inverter). A reduction-gear unit adapts the high rotational speed of the electrical machine to the wheel speed. A single-stage reduction is sufficient depending on requirements; where necessary, transmissions with two or more switchable reductions are used. The torque requested by the driver for propulsion via the accelerator pedal is sensed and processed by the vehicle control unit (VCU) (see Control of an electric drive). The VCU activates the inverter in accordance with the driver's request. The 12V vehicle electrical system also required for electric vehicles (or 24V electrical system for commercial vehicles) is supplied via a DC/DC converter by the HV electrical system (high voltage) (see DC/DC converter).

Regenerative braking system
The electric drive also incorporates a regenerative braking system in order to convert into heat during deceleration operations as little kinetic energy as possible via the service brake (friction brake, brake pad on brake disk). The aim with an electric drive is to keep the electrical machine in generator mode for deceleration operations and to store the power generated in the process in the battery (recuperation). The vehicle is decelerated in the process. The service brake is only then activated if the electrical machine's braking power

Figure 2: Components of an electric drive (schematic representation)
1 Electrical machine (e-machine),
2 Reduction-gear unit (single- or multi-stage),
3 Traction battery (HV battery),
4 Battery-management system,
5 DC/DC converter,
6 Power electronics (inverter),
7 Vehicle control unit,
8 Regenerative braking system,
9 Charging socket.



is not sufficient for the desired deceleration (see Regenerative braking system). The service brake is thus simultaneously protected. The stored energy can be used for propulsion.

Hybrid drive

Hybrid drives combine drives with combustion engine and electrical machine with suitably greater component outlay compared with a drive with combustion engine. There are different solutions, depending on the degree of electrification (see Figure 1).

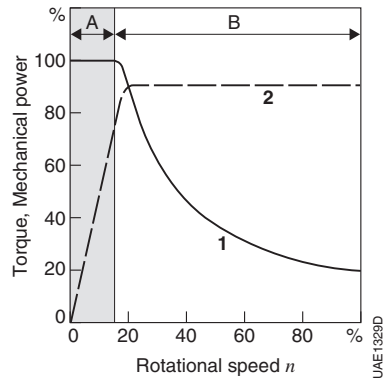
Because power is additionally made available by the electrical machine, the combustion engine can be designed with smaller dimensions (downsizing) to obtain the same overall power.

Outlay can also be reduced in the transmission. The power take-up element can be dispensed with when the electrical machine is used for starting. Reversing (backing up) can be effected electrically so that the reverse gear can be dispensed with in the transmission (see Transmission with integrated electrical machine).

The dimensioning of the components for the electric drive in the network of a hybrid drive – electrical machine, battery, and inverter – depends on the degree of electrification. In particular the costs of the battery with a voltage of up to 400 V in a full hybrid are substantially higher than the costs of a 48V battery in a mild hybrid.

Figure 3: Basic torque and power curves of an electrical machine

1 Torque,
2 Mechanical power,
A Base-speed range,
B Field-weakening range.



Advantages of an electric drive in an electric vehicle

Torque curve of electrical machine

Unlike the combustion engine an electrical machine already outputs its maximum torque from a rotational speed of zero (Figure 3, see also Electrical machines for drives). The rotational speed when driving is so high that it must be adapted to the required wheel speed via a reduction-gear unit. The torque transmitted to the wheels increases in the same proportion.

Transmission

The high torque available over a wide rotational-speed range enables the vehicle to be driven over a wide driving-speed range with only one transmission reduction. Single-stage transmissions are primarily used for passenger cars. Two-speed or multi-speed transmissions are also used for sports cars which reach higher speeds and for commercial vehicles. A reverse gear is not required in the electric drive's transmission because the electrical machine changes the direction of rotation when activated accordingly. Consequently, the transmission for an electric drive is much simpler and compact in terms of design than the transmission of a drive system with combustion engine.

Efficiency

For a combustion engine the maximum efficiency is around 40 % in the low speed range at high load. A conventional route profile however requires part load in the event of poorer efficiency.

Electric drives operate independently of the operating point, i.e., even with a low load request, with an efficiency of over 90 %.

Energy consumption

Because the efficiency of an electrical machine is significantly better than that of a combustion engine, energy consumption is lower for electric drives in an electric vehicle.

Zero local emissions
The electric drive has zero local emissions. The charging current required to charge the battery does however, depending on how the current is generated, cause emissions at the point of origin. CO_2 emissions are created if the current is produced in gas- or coal-fired power plants. Only if current from regenerative sources is used (e.g., via solar cells or wind turbines) is the electric drive free from CO_2 emissions.

Operating noises
The operating noises of an electrical machine are low such that an electric vehicle is very quiet at low speeds where the other driving noises (rolling noise of the tires, aerodynamic noises caused by the air stream) are not yet pronounced.

Vibrations
Due to the rotational movement of the electrical machine, hardly any vibrations are generated in the drive in an electric vehicle.

Component outlay
The combustion engine with its extensive peripherals (e.g., air-intake system, exhaust system) requires a lot of installation space. The components of an electric drive – with the exception of the traction battery – require much less space.

Advantages of a hybrid drive
The electrical machine in a hybrid vehicle improves the utilization of the combustion engine compared with a drive solely with a combustion engine thanks to the following operating modes.
– Recuperative braking: The braking effect and thus the recuperated brake energy depend however on the electrical machine's power.
– Boosting: The electrical machine can assist the combustion engine in the hybrid vehicle.
– The combustion engine can be operated at a higher load point with greater efficiency (load-point shift), the energy required for propulsion is stored via the electrical machine in the battery and can be utilized later for propulsion. This improves fuel consumption. The amount of fuel that can be saved depends on the degree of hybridization.
– The combustion engine can be switched off when the vehicle is stationary (start/stop system).
– At low speeds and with only a small power demand (e.g., in urban traffic) a full-hybrid vehicle with electric drive and switched-off combustion engine can drive in parts with zero emissions. The average fuel consumption of the combustion engine is lower in combination with the electric drive; as a result the CO_2 emissions are also reduced.
– Starting operations can be performed with the electric drive; the combustion engine can be cut in from a specific speed while the vehicle is rolling. This delivers a quiet, low-emission starting operation.

Compared with an electric vehicle the hybrid vehicle has the advantage of high range, comparable with that of a conventional vehicle with a combustion engine.

Electrified ancillary systems

The mechanical power dissipation of a drive system with internal-combustion engine includes, as well as the engine friction losses, the outlay on driving the accessories. These include, for example, the coolant pump, the engine-oil pump, the alternator, the high-pressure pump for fuel injection, the brake vacuum pump, the air-conditioning compressor, and the power-steering pump for the hydraulic power-assisted steering system. These accessories are typically driven for the most part by the internal-combustion engine via V-belts or toothed belts (Figure 4a). The internal-combustion engine must cater for the power demand of these accessories and also overcome the losses in the drive belt. This entails additional fuel consumption.

Demand control of the accessories

The power consumption of the driven accessories can be reduced by demand-oriented control. One example of this is intelligent alternator regulation. By specifying the control voltage it is possible to specifically control the power consumption. The alternator is still however driven via the lossy belt drive.

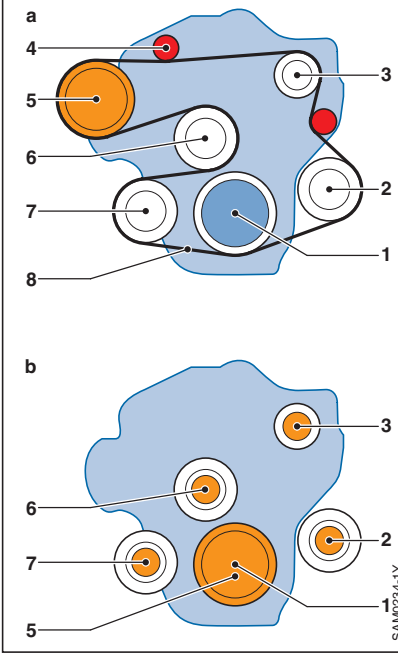
Electrically driven accessories

The accessories are driven even when they are not required. One example of this is the power-steering pump. When driving straight-ahead on expressways/freeways steering assistance is only needed on rare occasions, but the power-steering pump is permanently driven and causes losses. Electric power-assisted steering, which only cuts in when steering assistance is required, has in the meantime gained acceptance for use in passenger-car steering systems. The motivation behind these systems was however the possibility of realizing automated parking maneuvers.

Electrical driven accessories are no longer driven by the internal-combustion engine (Figure 4b) and can thus be operated in terms of their speed independently of the engine speed. The speed and thus the power consumption of the accessories are adapted to the demand. As the speed reduces, so too do the friction losses in the accessories. Coolant and lubricant flows can thus be adapted to the load and speed of the internal-combustion engine, which also impacts positively on the pollutant emissions.

Another advantage of electrified accessories is that they can even be operated while the engine is at a standstill if their operation is required during the standstill phase.

Figure 4: Accessories on an internal-combustion engine (schematic diagram)
a) Belt-driven accessories,
b) Electrified accessories.
1 Crankshaft, 2 Air-conditioning compressor,
3 Mechanically driven supercharger,
4 Tensioning system, 5 Alternator,
6 Coolant pump, 7 Power-steering pump,
8 Toothed belt,
9 Electrical machine on crankshaft.



Regenerative braking system

In the case of regenerative braking (also known as recuperation), the vehicle's kinetic energy during deceleration is converted by the electrical machine, which is operated as a generator for this purpose, into electrical energy. In this way, some of the energy which is normally lost as frictional heat during braking is fed in the form of electrical energy to the battery and then utilized.

Drag-torque simulation
A simple way of realizing regenerative braking is drag-torque simulation. Here the electrical machine is operated as a generator as soon as the driver releases the accelerator. The driver does not need to press the brake pedal for this purpose. In the case of a full hybrid, the internal-combustion engine is decoupled and the electrical machine exerts a generator torque of the magnitude of the engine's drag torque.

If the engine cannot be decoupled (as in the case of a mild hybrid), alternatively a lower generator torque can be exerted on the drivetrain in addition to the engine drag torque (drag-torque increase). The vehicle behavior changes only slightly compared with a non-hybrid vehicle.

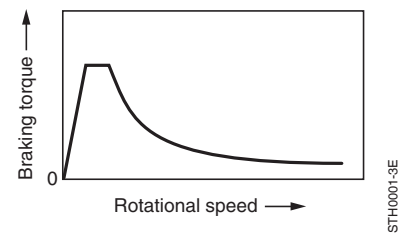
Simple regenerative braking system
During braking the electrical machine can exert an additional generator torque as well as drag-torque simulation and increase. The vehicle decelerates with the brake pedal in the same position faster than a comparable conventional vehicle. The available generator torque depends on the driving speed, the selected gear, and the state of battery charge. Therefore the vehicle can be subject to different levels of braking response, even with the brake pedal in the same position. This difference in braking response is perceived by the driver to be all the more disruptive, the greater the generator torque is part of the vehicle deceleration. For this reason, only small power levels can be recuperated with this simple regenerative braking system.

Cooperatively regenerative braking system
The service-brake system must be modified in the case of higher decelerations in order to further exploit the kinetic energy. To this end all or some of the service-brake friction torque must be replaced with a generator braking torque without the vehicle deceleration changing when the brake-pedal position and brake-pedal force are kept constant. This is realized in a cooperatively regenerative braking system, where the vehicle control and the braking system interact in such a way that constantly as much friction braking torque is recovered as generator braking torque can be replaced by the electrical machine.

During a braking operation the maximum generator torque that can be achieved by the electrical machine changes continuously over a wide speed range (Figure 5). This is due to the fact that the electrical-machine power (i.e., torque times speed) is constant in this range at high rotational speed. Only at low speeds is there a range of constant maximum electrical-machine torque. If the speed continues to drop, the regenerative braking torque that can be achieved again drops to zero.

A constant torque is required at the wheels for a constant deceleration of the vehicle. If the electrical machine operated as a generator is exploited to its limit torque, the friction braking torque must be continuously reduced with decreasing

Figure 5: Maximum torque path for electric-machine generator mode



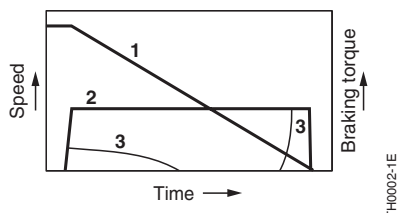
driving speed (and thus decreasing electrical-machine speed) because the generator torque increases. Figure 6 shows by way of example a braking operation of a power-split hybrid. At the start of the braking operation the generator torque is increased until it reaches its maximum level. If this corresponds to the demanded total braking torque, the friction braking torque can be reduced to zero. Towards the end of the braking operation the generator torque is reduced and replaced completely by a friction braking torque because the electrical machine can no longer deliver a generator torque at very low rotational speeds (Figure 5). During the braking operation from high driving speed to a standstill, therefore, the distribution between friction braking torque and regenerative braking torque is continuously adapted with constant pedal actuation.

ESC-based regenerative braking system

For the technically simplest implementation of a cooperatively regenerative braking system essentially the hardware of a conventional ESC system (Electronic Stability Control, electronic stability program) is used (Figure 7) and the operating strategy is enhanced by the recuperation functionality.

Basically, such a system can be operated both with an electromechanical and with a vacuum brake booster where the combination with an electromechanical brake booster offers additional advantages with regard to recuperation efficiency and redundancy for highly automated driving. The back pressure of the low-pressure reservoir can be reduced to further increase the recuperation efficiency. The exact relationship is described in more detail in the following.

Figure 6: Cooperative braking operation for a power-split hybrid
1 Driving speed,
2 Total braking torque,
3 Friction braking torque.



Operating principle

A pedal-travel sensor is additionally required to detect the driver's braking request. This sensor either is automatically made available in combination with an electromechanical brake booster or must be integrated in the system in combination with a vacuum brake booster by way of an additional external pedal-travel sensor.

The driver's braking request must be detected via the pedal-travel sensor because the hydraulic pressure built up in the braking system during recuperation generally cannot be clearly assigned to a specific pedal position.

The requested braking torque is initially made available by the electrical machine operating as a generator. If the generator torque is not sufficient, additional hydraulic pressure and hence a friction braking torque are built up at the basic brake. This process is called torque blending. When the pedal is pressed, the same amount of brake fluid as used in a purely hydraulic braking operation is displaced in the braking system. To achieve a reduced pressure or no pressure in the brake calipers, the excess brake fluid is intermediately stored in the low-pressure reservoir; this is called volume blending. The brake fluid is displaced via the inlet or outlet valves on the rear or front axle into the low-pressure reservoir (Figure 7). Through this, during the entire braking operation the back pressure of the reservoir chambers is applied at the respective wheel brake and results in the case of purely recuperative braking operations in an unwanted residual pressure and consequently also in a possible residual slip torque at the brake disks. Through a reduction of the back pressure and a high selected wheel-brake response pressure these systems can, despite the residual pressure, achieve recuperation efficiency of up to 100 % (recuperated energy com-

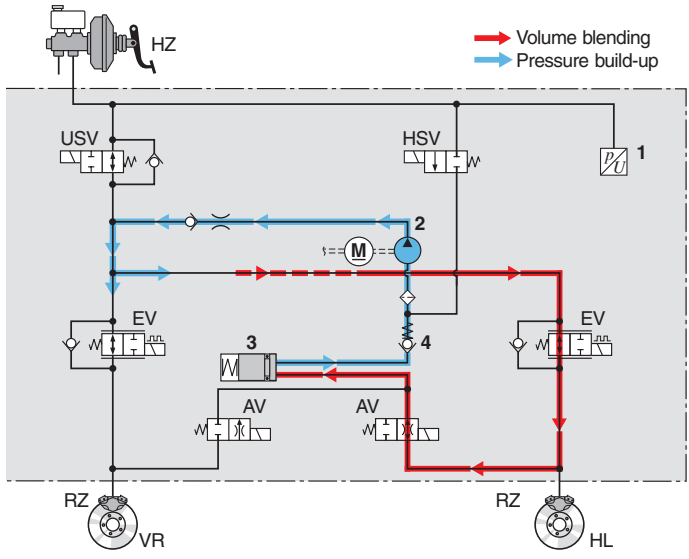
pared with the energy made available by the generator). If additional pressure is requested during the braking operation, brake fluid is taken from the low-pressure reservoir by means of the return pump and displaced into the wheel brakes (Figure 7).

During recuperation the hydraulic pressure in the braking system is as a rule below the expected pressure of a purely hydraulic braking operation. If the brake booster has not yet exceeded the point at which the brake pedal is hydraulically coupled with the braking system, the pressure reduction is not yet detected at the brake pedal. However, as soon as this point is overcome, the original pressure must in the case of a vacuum brake booster be built up at least on one axle in order to ensure an unchanged pedal force and a consistent pedal feel. On the other hand, when an electromechanical brake booster is used, the missing pressure can be compensated for by reducing the electromechanical assisting force so as not to experience any limitation in recuperation or any reduction in recuperation efficiency. The compensation of the missing system for an unchanged pedal feel is also called force blending.

Highly automated driving

The integration of regenerative braking systems suggests itself particularly for vehicles with highly automated driving functions because these vehicles are as a rule electrically chargeable hybrids or even purely electrically powered vehicles. In the case of highly automated driving functions (level > 2) the combination with an electromechanical brake booster additionally provides the required redundancy

Figure 7: Hydraulic unit of an ESC-based regenerative braking system
(Depiction of one of the two brake circuits.)
1 Pressure sensor, 2 Return pump with pump motor (M), 3 Low-pressure reservoir, 4 Non-return valve.
HZ Brake master cylinder, RZ Wheel-brake cylinder, HSV High-pressure switching valve, USV Switchover valve, EV Inlet valve, AV Outlet valve, V Front, H Rear, R Right, L Left.



in order to safely brake the vehicle fully automatically to a standstill in the event of the total failure of a component.

Braking-force distribution
As with the design of conventional brake systems, so too with the design of a regenerative brake system the braking-force distribution between front and rear axles is of crucial importance to the vehicle's directional stability. As deceleration increases, the normal force on the front wheels increases while the normal force on the rear wheels decreases. If the electrical machine is connected to the front wheels, a greater wheel braking torque can also be transmitted with increasing deceleration and thus increasing normal force on the front axle. On a rear-wheel-driven vehicle or a vehicle with an electrified rear axle (i.e., internal-combustion engine on the front axle and electrical machine on the rear axle) the transmittable regenerative torque decreases with increasing deceleration.

The electric power of the battery also has a major influence on the utilization of recuperation since it is the limiting element for absorbing electrical energy from the vehicle movement. The maximum possible purely recuperative deceleration increases with increasing energy-accumulator power. For directional-stability reasons, however, a large generator braking torque can only be transmitted on the front axle.

Influence on driving-dynamics control
Since the regenerative brake system influences braking stability, the driving-dynamics control must be adapted to the changed physics of driving. It is expedient to suppress the regenerative component of braking when unstable driving states or excessively high brake slip are detected and to realize deceleration and stabilization interventions solely by means of the friction-braking system. Otherwise instabilities and drivetrain vibrations could disrupt optimum wheel-slip control.

In relation to partial-braking operations interventions by the vehicle stabilization system are so rare that the suppression of recuperative braking in these situations has no noticeable impact on the average energy consumption for an electrified vehicle.

Thermal management for electric vehicles

Thermal management refers to the control of heat flows in a vehicle. The function of thermal management is to protect components against overheating, operate these components in their optimal temperature ranges, and generate pleasant temperatures for the occupants in the passenger compartment. The thermal-management system switches into the appropriate operating modes in response to the heating or cooling demand of the components or the passenger compartment. The outside temperature is taken into consideration in the process. The correct configuration of thermal management is becoming ever more important and provides for increased range and comfort in electric vehicles.

Intelligent thermal management results in

– increased range for electric vehicles,
– full air-conditioning comfort through the targeted use of heat and cold,
– rapid charging by efficient battery cooling – even at high temperatures,
– longer battery service life due to lower thermal load.

Operating principle

A thermal-management system in an electric vehicle is generally more complex than in conventional vehicles with internal-Internal-combustion engines. Thus the e-axle (integration of electrical machine, power electronics, and transmission, see Electric final-drive unit) is constantly cooled while the battery must be either cooled or heated depending on the situation (Figure 8). Furthermore, the waste heat from an internal-combustion engine is no longer available to heat the passenger compartment, and consequently energy-efficient measures, e.g., a heat pump, are used for this purpose.

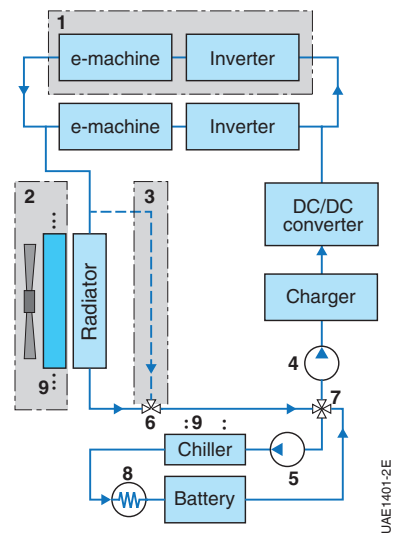
The refrigerant circuit and cooling circuit must come together in optimal interaction to transfer the heat in the vehicle and to make available the necessary temperatures. The routing of the two circuits changes depending on the heating or cooling demand. This gives rise to different operating modes.

Cooling circuit

In the cooling circuit coolant is circulated by an electric coolant pump to deliver a coolant flow in line with demand. The coolant transfers heat from where it is generated to the point in the vehicle where it is needed. Coolant switching valves direct the coolant flow in the secondary cooling circuits of the drivetrain and in the heating circuits of the passenger compartment for this purpose. Thanks to the high specific heat capacity the coolant can absorb a lot of heat in a very small space, which is necessary for example to cool the e-axle or the battery effectively.

Likewise, heat can be distributed very flexibly in the vehicle with the coolant. When coolant absorbs heat, its temperature increases and it must be cooled in a heat exchanger.

Figure 8: Cooling circuit in an electric drive
Refrigerant circuit is not pictured here.
1 Second electric drive (optional),
2 Fan and condenser,
3 Radiator bypass (optional),
4 Coolant pump,
5 Coolant pump,
6 Switching valve,
7 Switching valve,
8 Battery heater,
9 Connection to refrigerant circuit.



Refrigerant circuit
In the refrigerant circuit a refrigerant, which can be both liquid and gaseous, circulates. When the refrigerant evaporates (transition from liquid to gaseous) a refrigerating capacity is generated which facilitates cooling even below the ambient temperature. This established principle of air-conditioning the passenger compartment in the summer is also used to cool the battery at very high outside temperatures. The heat released during condensation (transition from gaseous to liquid) can furthermore be used in the winter to heat the passenger compartment. The refrigerant circuit is driven by an electric air-conditioning compressor, which compresses the refrigerant to the desired pressure so that evaporation and condensation take place at the desired temperature in each case.

Driving in the winter
At low outside temperatures both the passenger compartment and the battery must be heated in order to ensure full drive power at all times. The passenger compartment is heated by the refrigerant circuit by way of a heat pump. The battery is heated by an electric flow-type heater and is additionally supported with waste heat from the e-axle.

The use of a heat pump improves the efficiency of the passenger-compartment heating, which has a positive effect on the range. The use of waste heat from the e-axle to heat the battery both improves efficiency and ensures that the battery is fully available at all times.

Driving in the spring and fall
At moderate outside temperatures the battery and e-axle can be jointly cooled via a coolant circuit, whereby the waste heat is dissipated via the radiator to atmosphere. As a result, the energy consumption for thermal management is reduced to a minimum, which impacts positively on efficiency and range. At the same time the battery and e-axle can be set to optimal temperature levels which enable the drive to operate at optimal efficiency.

Driving in the summer
At high outside temperatures the battery must be cooled via the “chiller” (heat exchanger between the cooling and refrigerant circuits) by the refrigerant circuit in order to guarantee the maximum permissible temperature in the battery and thereby ensure a long battery service life.

The refrigerant circuit is likewise responsible for air-conditioning the passenger compartment. The e-axle is cooled separately from the battery via the radiator, facilitating unlimited, energy-efficient cooling of the e-axle.

Charging
As a rule limited charging capacity with alternating voltage is available at home or at the workplace for charging the battery. The charger, which is responsible for converting the alternating voltage into direct voltage, and the battery are jointly cooled via a cooling circuit and the radiator. In this way the charging losses can be kept as low as possible, which has a positive impact on the battery-charging time.

During rapid charging at high outside temperatures the temperature in the battery increases very quickly. So as not to exceed the maximum permissible temperature and to enable the heat to dissipate in a short time, the battery is cooled via the “chiller” by the refrigerant circuit. This ensures a rapid charging process at all times.

Charging infrastructure for electrified vehicles

Facts
Electrical energy today is very widely used, very reliably available, and in every home and business at that. It seems reasonable to charge electric vehicles from the already existing mains-supply network as well.

Because the amounts of energy and in part also the desired charging capacities are a matter of some significance, an electric vehicle cannot be considered as an electrical domestic or industrial appliance. Charging sockets for vehicles therefore very often have special designs.

The sockets for the consumers make available worldwide alternating current (AC) in all major mains-supply networks. In some cases this may be three-phase alternating current. All batteries require direct current (DC) for charging. Consequently, charging vehicle batteries necessitates a battery charger that has two functions:
– converting from alternating to direct current and
– at the same time the charger must, depending on the current state of the battery, be able to adjust current and voltage individually.

Charging types
The charger can be accommodated in the vehicle ("on-board charger"). In this case the process involves alternating-current charging (AC charging) because the vehicle is supplied with alternating current (see On-board charger).

Alternatively the charger can also be accommodated in the infrastructure, for example in the charging post. In this case the process involves direct-current charging (DC charging) because the vehicle is supplied with direct current. It must still be borne in mind that vehicles cannot be connected to any direct-current sources because the charger is not just a rectifier.

Both charging types have their justifications. The charger carried in the vehicle can be connected virtually anywhere to existing alternating-current networks, but is generally limited in its charging capacity by installation space, weight, and costs. Furthermore, large capacities are only sensible to a limited extent in that larger capacities cannot be drawn from the mains-supply network everywhere.

When the charger is integrated in the infrastructure, this opens up the possibility of very high charging capacities because as a rule weight and installation space are less critical than in the vehicle itself. Sometimes the increased costs can be apportioned to a number of vehicles which use this charger one after the other. But the costs do not grow linearly with the capacity of the charger because upstream components of the mains-supply network must "grow alongside". Apportioning the additional costs for high charging capacities will accordingly always mean an increased price per energy (kWh), which must ultimately be billed.

A further additional expense associated with direct-current charging is appropriate signal transmission in that only the vehicle knows the battery condition precisely and consequently must control the charger in the infrastructure. Despite the expense, direct-current charging is today the method of choice for realizing very high charging capacities.

Orders of magnitude
To convey the orders of magnitude of the electrical engineering of vehicles, there follow some figures which are compared with households (in Germany).

Vehicle
As at 2020 figures of 13 kWh through 27 kWh per 100 km (with outliers up to over 40 kWh) are given for the energy demand per distance traveled for passenger cars. This is naturally also usage-dependent (dependent on the driving style) and additionally – in contrast

to internal-combustion engines – clearly weather-dependent. Heating (in the winter) and cooling (in the summer) mean a marked increased demand with regard to electric vehicles. 20 kWh per 100 km can be envisaged as an average. The other way round, this means: the driver can drive 5 km on 1 kWh of electrical energy. This figure may change somewhat over the years because the current trend is towards designing economical vehicles because economy with restricted battery capacity means increased range.

In the future larger and more cost-effective batteries as well as more luxurious vehicles can presumably be envisaged. Assuming that an average passenger car covers 10,000 km per year (naturally this value is subject to significant spread), a vehicle with the above-mentioned average value estimated over a year requires approximately 2,000 kWh.

Household

By way of comparison, the average annual consumption for households is in the order of 3,500 kWh per year – here too there is a very wide spread of around 500 kWh through 25,000 kWh per year – this would be similar to the spread of the annual mileage of passenger cars. In other words, in terms of energy a passenger car is to be considered as a small household.

Charging capacity

The sizes of the batteries used in vehicles cover a considerable range. The cause is a certain dilemma in the vehicle design because the weight and costs of the batteries are very high and can also make up a significant proportion of the weight and costs of the complete vehicle. In battery-powered electric vehicles battery capacities of 16 kWh (for a subcompact-size car) to over 100 kWh (for a luxury-class vehicle) may be given by way of example. The amount of energy that is charged to the vehicle in a single charging process is smaller. How much smaller depends very much on the user because the battery can be recharged for any partial states of charge and is not necessarily fully charged. Which charging capacity is required to recharge the mentioned energy also depends on the desired charging time.

Plug connectors

There are very many different socket-outlets used in public mains-supply networks. This is not only due to national-market variants, but also due to their purpose and power capacity. The difference between domestic socket-outlet (“Schuko” grounding-type) and industrial socket-outlet as per IEC 60309 [2] (“Cekon”) is mentioned by way of example. Continuous power of approximately 2.3 kW and short-time power of 3.7 kW can be drawn from a domestic socket-outlet. The figures can be lower due to other power consumers which are operated at the same time. Continuous power of 3.7 kW to over 100 kW can be drawn from an industrial socket-outlet (depending on the size). The types for 11 kW or 22 kW are very common.

To charge vehicles, a purely mechanical adaptation (as familiar from “travel plugs”) is conceivable but not sensible because the charging capacity would have to be designed for the weakest adaptable socket-outlet. However, there is often the wish to utilize the available current source or use the maximum capacity of the charger in the vehicle.

For this reason specific plug connectors have been introduced for electric vehicles which contain, in addition to the contacts for the charging current, signal contacts which communicate to the vehicle the load capability of the current charging source.

References

[1] https://wirtschaftslexikon.gabler.de/
[2] IEC 60309: Plugs, socket-outlets and couplers for industrial purposes.

Control of electric drives

Drivetrain of an electric vehicle

In an electric vehicle propulsion is provided by an electrical machine (e-machine). The electrical energy required for this is stored in battery-powered vehicles in the traction battery; in fuel-cell vehicles the energy is continually generated in the fuel cell. Direct current is available in both cases. However, a three-phase alternating current is required to operate the e-machine. Direct current is converted to alternating current in the inverter.

These components must be activated in a coordinated fashion in order to convert the torque request input by the driver via the accelerator pedal into an e-machine torque. The torque then provides the drive or decelerates the vehicle. All these operations are controlled in an electronic control unit called the VCU (Vehicle Control Unit).

Figure 1 provides an overview of the components in the drivetrain of a battery-powered vehicle.

Vehicle control unit for electric vehicles

Function

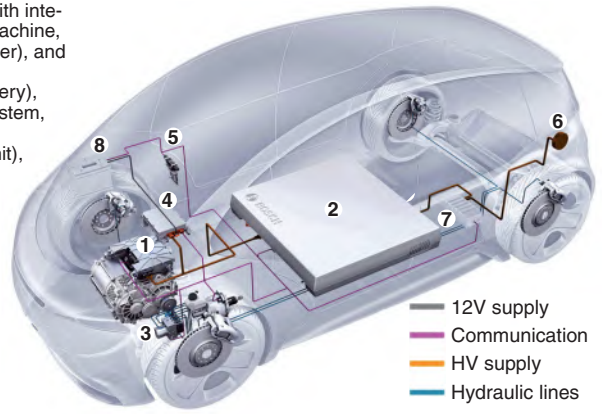
The electronic control of electric vehicles (vehicle control unit, VCU) is responsible for converting the driver command into propulsion. The drive torque required for this is relayed to the inverter control unit to implement e-machine drive power. Furthermore, the subsystems of the HV vehicle electrical system (HV, High Voltage) and the components for heating and cooling the cooling circuits in the vehicle are coordinated.

Electrical system overview

Figure 2 describes the ECUs and subcomponents in the drivetrain of an electric vehicle. Aside from the VCU as the central processor there is among others the battery control unit, which is responsible for opening and closing the contactors in the traction battery. The charging control unit is responsible for coordinating the charging process, firstly for charging with alternating current and secondly for rapid charging with direct current at charging posts. The inverter control unit regulates the current through the e-machine. The HV consumers for heating and cooling (electric continuous-flow heater and

Figure 1: Components in the drivetrain of a battery-powered vehicle

1 Electric final-drive unit with integrated components: e-machine, power electronics (inverter), and transmission,
2 Traction battery (HV battery),
3 Regenerative braking system,
4 DC/DC converter,
5 VCU (Vehicle Control Unit),
6 Charging socket,
7 On-board charger,
8 12V battery.



air-conditioning compressor) are likewise featured and are activated and deactivated by the VCU. The DC/DC converter supplies the LV vehicle electrical system (Low Voltage,12 V for passenger cars and 24 V for commercial vehicles) with electrical energy. This vehicle electrical system contains the consumers such as for example the infotainment and further ECUs in the vehicle.

Functional system overview

A VCU has many other subtasks aside from the primary tasks of implementing the driver command, coordinating the subsystems of the HV vehicle electrical system, and controlling the cooling circuits in the vehicle.

Torque Demand subsystem

The driver issues a concrete driver command through the accelerator-pedal position. The accelerator-pedal position is converted into a setpoint value for the drive or deceleration torque.

If the accelerator-pedal position is in a low range (e.g., 0 to 15 %), a negative torque request is made (Figure 3). The e-machine is in this case used to decelerate the vehicle and the electrical energy generated by recuperation is returned to the traction battery. A medium accelerator-pedal posi-

Figure 3: Torque request with one-pedal control

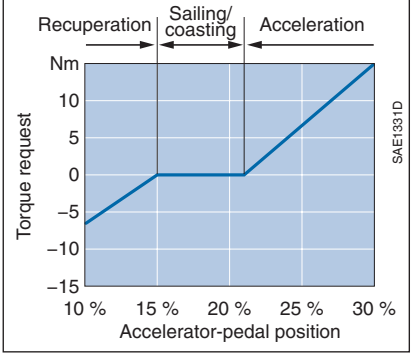
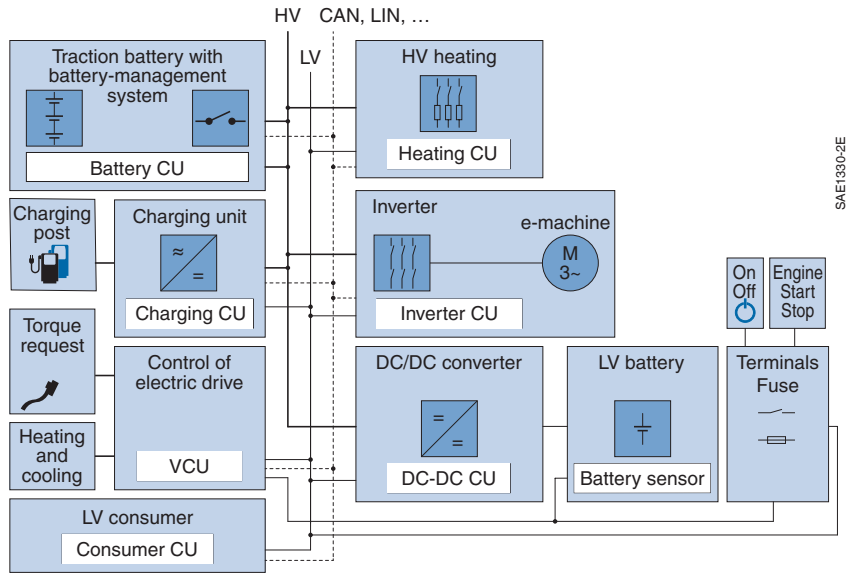


Figure 2: Electrical system diagram for an electric vehicle

HV High Voltage, LV Low Voltage (12 V for passenger cars, 24 V for commercial vehicles), AC Alternating Current, DC Direct Current, ECU Electronic Control Unit.



tion (e.g., from 15 to 21 %) is interpreted in such a way that driving at constant speed is desired. The ehicle behavior here is similar to the sailing (coasting) mode of a vehicle with an internal-combustion engine. A high accelerator-pedal position (e.g., greater than 21 %) requests a pedal-position-dependent acceleration. Acceleration and braking with the accelerator pedal can be controlled by means of this one-pedal control.

The vehicle's driveability is influenced by selecting a driving profile – Normal, Eco or Sport. In the Eco driving program a lower drive torque is requested with the same accelerator-pedal position (Figure 4). More drive torque is requested when the Sport driving program is selected.

In addition, there is an interface through which the driver can adjust the recuperation level (e.g., selector-lever position “B” or an additional switch). The higher the recuperation level, the higher the requested recuperation torque. Figure 5 describes the torque requests for recuperative braking at different recuperation levels.

As well as selector-lever position “B”, by means of which the recuperation mode is activated, there are usually still further selector-level positions known by vehicles with automatic transmissions:

– P (Park): the vehicle is parked and secured against rolling away.
– N (Neutral): the drivetrain has no frictional connection. The vehicle is not secured against rolling away.
– D (Drive): the vehicle is in the driving mode in the forward direction.
– R (Reverse): the vehicle is in the driving mode in the reverse/backing-up direction.

Aside from the direct torque input the driver can make an indirect torque input via the cruise control (vehicle-speed controller) (Figure 6). A setpoint drive torque is calculated, depending on the current driving condition. A speed limiter restricts the maximum speed to the value set by the driver. The creeping function is activated at low speeds when the vehicle is in the driving mode and the accelerator pedal is not pressed. The vehicle then moves at a low speed (e.g., 5 km/h). This simplifies vehicle handling in certain situations, for example when parking.

Unlike vehicles with internal-combustion engines, there is no idle speed in electric vehicles when the accelerator pedal is not pressed and the vehicle is stationary.

Figure 4: Torque request with one-pedal control for different driving programs
(Sailing/coasting range shown applies to Eco driving program.)
1 Eco driving program,
2 Normal driving program,
3 Sport driving program.

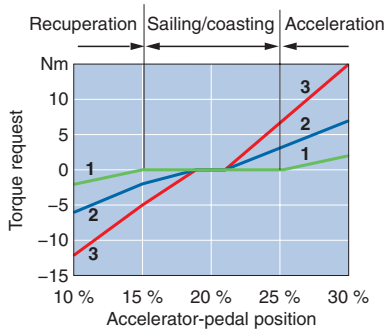
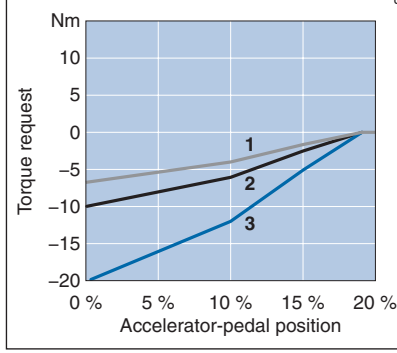


Figure 5: Torque request for recuperative braking for different recuperation levels
1 Recuperation level 1,
2 Recuperation level 2,
3 Recuperation level 3.



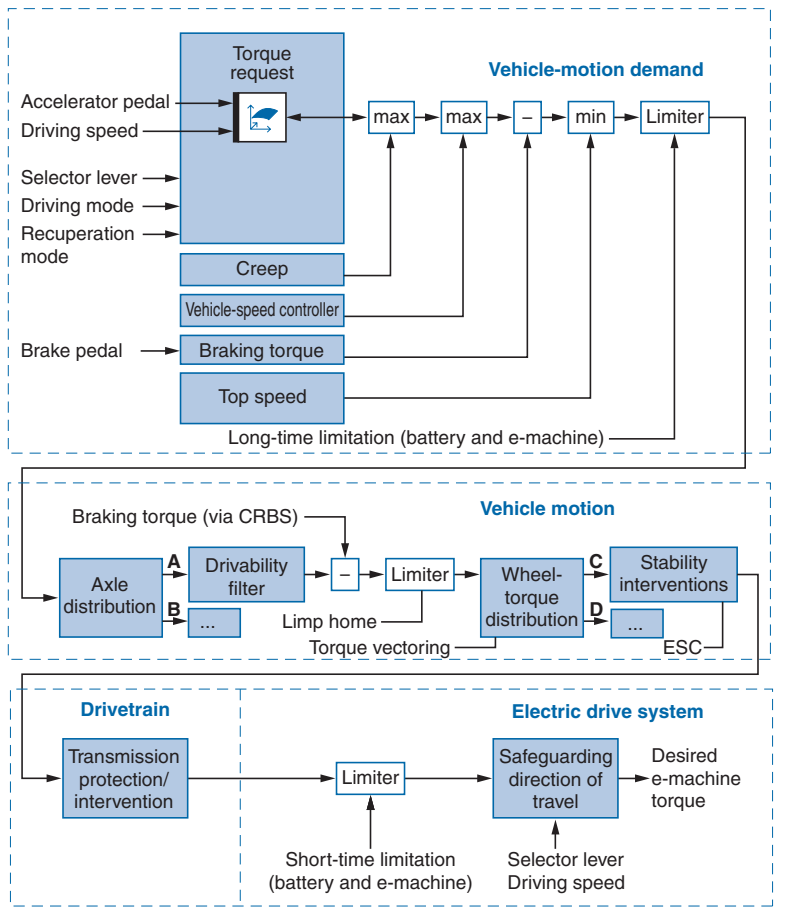
In vehicles of some manufacturers one-pedal control is used to bring about deceleration via the e-machine but also activation of the mechanical friction brake. In some countries there are standards which provide for an activation of the stop lamps from a certain deceleration. Despite the possibility of activating the mechanical friction brake via one-pedal control there is however still a brake pedal, which facilitates safe and direct activation of the friction brake.

Torque Structure subsystem

The ECU for driving-dynamics control (ESC, Electronic Stability Control) spe-

Figure 6: Coordination of torque request (Torque Structure)

A Path for first e-axle,
B Path for second e-axle (corresponds to path A),
C Wheel-torque distribution for left wheel,
D Wheel-torque distribution for right wheel (like path C).
ESC Electronic Stability Control (electronic stability program),
CRBS Cooperative Regenerative Brake System,
Torque Vectoring: Wheel-individual drive-torque distribution.



cifically initiates brake interventions in order to guarantee vehicle stability. In the Torque Structure subsystem the individual torque requests from the Torque Demand subsystem, the brake intervention by the ESC, and the request for purely generative deceleration by the regenerative braking system are coordinated. The functionality contains a filter for smoothing the torque request and thus for improving driveability and avoiding load reversals. The torque calculation takes into account the position of the selector lever and thus it is ensured that the vehicle moves in the forward direction in the "Drive" position and in the reverse direction in the "Reverse" position. No electrical torque is generated in the "Neutral" and "Park" selector-lever positions (Figure 6).

Electrical Supply System subsystem
A further task of the VCU is to coordinate the components of the HV vehicle electrical system. When the driver requests a vehicle drive-ready state (e.g., via the start button), the VCU sends a request via the CAN bus to the battery control unit to close the contactors in the traction battery. Once the HV vehicle electrical system has been successfully activated, the VCU sends a further request to the ECU of the DC/DC converter so that the latter starts with the voltage conversion from the high-voltage side to the low-voltage side (12V side for passenger cars or 24V side for commercial vehicles). This ensures that the low-voltage supply is redundant.

Furthermore, HV consumers (e.g., the PTC booster heater or the HV air conditioner) can be deactivated by the energy management where necessary.

Electric vehicles also have a 12V or 24V battery because this is needed during the first starting operation to check the HV system.

Thermal Management subsystem
In the Thermal Management subsystem the VCU controls the components for heating and cooling the cooling circuits in the vehicle. Typically there is a cooling circuit for the drive, in which the e-machine, the inverter, the DC/DC converter, and the charger are located. Depending on the temperatures of these components the VCU activates the coolant pump in the cooling circuit and this pump cools the specified components accordingly. The heat transferred to the coolant is dissipated via the radiator to atmosphere. In the event of a higher cooling demand the radiator fan is additionally activated. The aim is to protect the components against overheating.

The radiator fan can, depending on the installation location of the traction battery, also be used to cool the battery. In the case of a separate cooling circuit for the battery the VCU controls a second coolant pump for this cooling circuit. If the battery temperature is very low (e.g., below 10 °C), the VCU activates the PTC booster heater in the battery circuit so that the traction battery is heated up. At very high outside temperatures (e.g., above 40 °C) it activates the air-conditioning compressor, which can lower the coolant temperature and with it also the battery temperature. This is necessary especially on hot days, at high loads during driving or when rapid charging because overheating the battery could shorten its service life.

Vehicle Functions subsystem
The VCU computes the remaining range that the vehicle can cover on the current battery charge. The computation is made on the basis of the available energy in the traction battery and the energy consumption in the vehicle. The remaining range is sent on the CAN bus to the display ECU so that the value can be seen by the driver in the display.

Further applications of the VCU for electric vehicles

Battery-electric buses and commercial vehicles

Battery-electric buses and commercial vehicles can be used for passenger transportation and goods deliveries in urban environments. The operating principle of the VCU in these applications is very similar to those for passenger cars.

Range-extender operation

There are electric vehicles which are equipped with an internal-combustion engine that operates as a range extender. An additional e-machine is operated by the combustion engine and generates electrical energy in generator mode. The power figures of the combustion engine and the additional e-machine are much smaller than the power of the electric driving motor. Because the combustion engine is mechanically independent of the drivetrain, it can operate in at an optimal operating point. The electric power generated in this way is stored in the traction battery or made available to the electric drive. As a result, the range of the electric vehicle is increased when compared with pure battery mode.

A fuel cell may also be used as a range extender instead of an internal-combustion engine. The operating principle of the VCU is in this case very similar.

Coordination of the range extender with the VCU

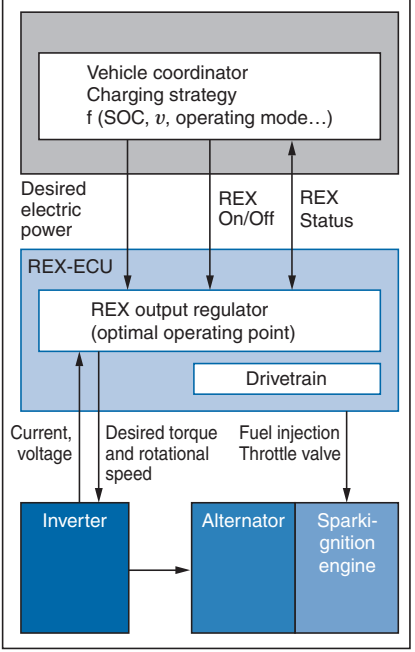
When the state of charge of the traction battery is low, the VCU issues a power request to the range extender's engine control unit (REX-ECU). The vehicle coordinator implemented in the VCU for range-extender operation determines the desired electric power (Figure 7) depending on the

– charging strategy,
– operating mode,
– power limitations of the battery,
– start/stop requests.

The engine control unit activates the range-extender internal-combustion engine by inputting variables for fuel injection and throttle-valve position and requests a torque via the additional inverter, which activates the e-machine operating in generator mode. The output regulator in the engine control unit computes the desired e-machine speed and the desired torque. The inverter coordinates the operating mode and sets the desired values (torque, speed, limitations). The requested electric power can thus be made available.

Figure 7: Coordination of the range extender with the vehicle control unit

REX Range Extender,
ECU Electronic Control Unit, SOC State of Charge.
v Driving speed,
f Function.



Hybrid drives

Motivation

A hybrid electric vehicle (HEV), also called a hybrid vehicle, has two or more energy storage devices with their associated energy converters which drive the vehicle jointly or separately (EU Directive 2007/46/EC, [1]). In most hybrid vehicles one or more electrical machines operate as well as the internal-combustion engine. The energy for driving is drawn from a fuel tank and a traction battery (Figure 1). The traction battery is based for the most part on lithium-ion technology with voltage levels of 48 to 800 V.

Essentially three objectives are pursued with the use of electric hybrid drives, where a vehicle driven purely by an internal-combustion engine serves as the reference:

– reduced fuel consumption and thus CO_2 emissions,
– reduced exhaust emissions (including zero emissions by electric driving), and
– increased torque and power output to improve operative dynamics.

Features

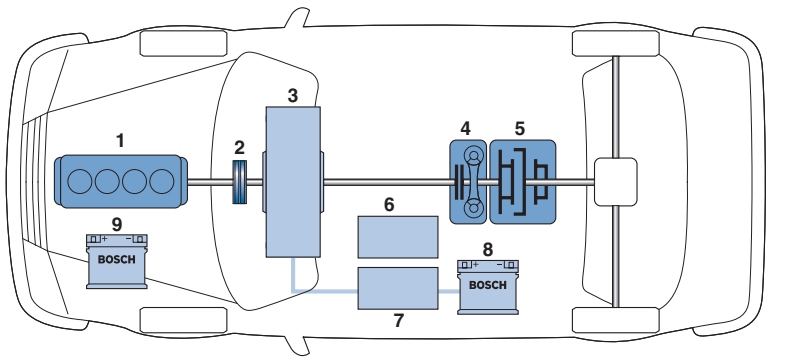
Hybrid vehicles combine many benefits of both types of drive in the drivetrain. The extended ranges of an internal-combustion engine and rapid refueling are combined with the high efficiency of an electrical machine also in urban environments at low driving speeds and with many stops at traffic lights.

The internal-combustion engine and the electrical machine can appear in the drivetrain in a variety of drive structures. Here vehicle transmissions (if necessary with one or more integrated electrical machines) also play a role, mostly to provide the internal-combustion engine with variable gear ratios for power conversion or to drive or decelerate the vehicle electrically.

The electrical machine offers high torques at low engine speeds. It thus complements the internal-combustion engine, whose torque only starts to increase at medium engine speeds. The electrical machine and the internal-combustion engine contribute to vehicle motion to varying degrees, depending on the drive power demanded.

Figure 1: Components of a hybrid drive

1 IC engine, 2 Clutch, 3 Electrical machine, 4 Power take-up element, 5 Transmission, 6 DC/DC converter, 7 Inverter, 8 Traction battery, 9 12 V battery.



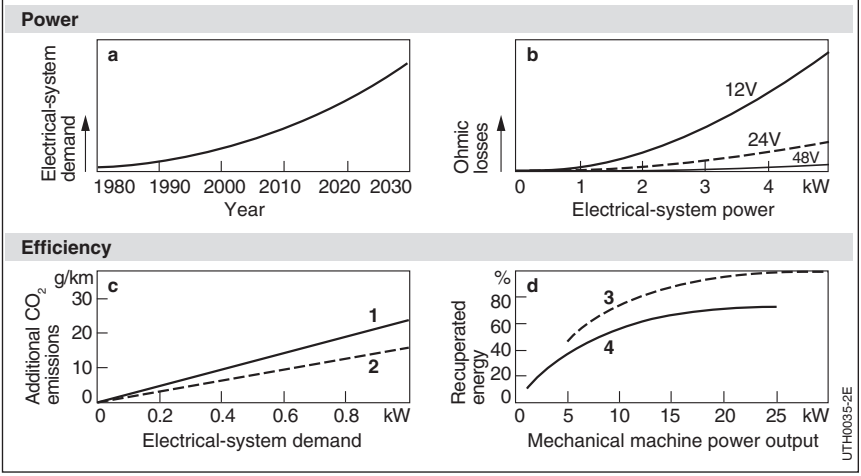
Reducing the electrical losses in the 12 V vehicle electrical system

The electrical power demand in passenger cars is rising continuously and will continue to increase in the next few years due to the use of an increasing number of powerful electrical comfort components (Figure 2a). The electrical energy required to supply the electrical loads/consumers in the 12 V vehicle electrical system is currently generated in conventional vehicles by an alternator which is connected via a belt to the internal-combustion engine's crankshaft. The load on the internal-combustion engine for generating current increases the fuel consumption. For a mid-size vehicle the fuel consumption and hence the CO_2 emissions at an electrical-system load of 500 W increases typically by approximately 0.5 *l* per 100 km; the CO_2 emissions increase accordingly (Figure 2c).

A reduction in the CO_2 emissions can be achieved by lowering the electrical losses in spite of increasing electrical-system load. This can be achieved by increasing the voltage in the vehicle electrical system to for example 48 V. Increasing the voltage enables the current to be reduced with the same power output and thereby reduce the ohmic losses (Figure 2b). This becomes apparent particularly in high-power consumers such as an electric air-conditioner compressor; most consumers (e.g., ECUs) are operated in the 12 V vehicle electrical system.

Figure 2: Development of electrical-system demand and its influence on CO_2 emissions [2]

a) Increase in electrical-system power (trend),
b) Reduction of losses (trend),
c) Influence on CO_2 emissions in the underlying driving cycle,
d) Share or recuperated energy.
1 New European Driving Cycle (NEDC, previous driving cycle),
2 Worldwide Light-duty Test Cycles (WLTC, driving cycle to be applied since 09/2018 for newly registered passenger cars, since 09/2019 for newly registered light commercial vehicles),
3 Excluding combustion-engine drag losses (engine is disconnected via a clutch from the electrical machine),
4 Including combustion-engine drag losses (engine is connected to the electrical machine).



Functionalities of hybrid drives

The way in which internal-combustion engine, electrical machine and energy storage device (battery) interact determines the possible hybrid functionalities. Hybrid-drive systems can be classified on the basis of the implemented functions and the proportion of electric power that goes towards the entire drive (Table 1).

Start/stop function

The average fuel consumption and thus the CO_2 emissions can be reduced by switching off the internal-combustion engine when this is not required for propulsion of the vehicle. The electrical energy required to supply the vehicle electrical system during this time is delivered by the battery.

The start/stop function switches the internal-combustion engine off temporarily without the driver having to turn off the ignition switch ("the ignition"). The engine is typically switched off when the vehicle is stationary or during constant driving if no energy is required. It is then restarted automatically as soon as the driver is ready to drive off, accelerate or brake.

This functionality is achieved in many vehicles already with the existing 12 V vehicle electrical system with manageable technical outlay.

Recuperation

One measure for reducing fuel consumption is aimed at converting the vehicle's kinetic energy into electrical energy (recuperation). With an electrical machine power output of approximately 12 kW already a large proportion (roughly 80 %, see Figure 2d) of the available kinetic energy can be recuperated and stored in the battery. Depending on the electrical-system load and the vehicle, even more energy can be recuperated than is needed for the electrical-system supply. In this case it is expedient to export this energy in the form of kinetic energy back to the crankshaft.

In the case of recuperation the electrical machine is used to convert kinetic energy into electrical energy. Braking operations are effected where possible not – or at least not only – by the friction torque of the service brake, but by a generative braking torque of the electrical machine (regenerative braking). The electrical energy is transferred to and stored in the battery (Figure 3).

Hybrid driving

Hybrid driving refers to those states in which the internal-combustion engine and the electrical machine are in opera-

Figure 3: Recuperation
The arrows indicate the energy flow.
1 IC engine,
2 Electrical machine,
3 Battery.

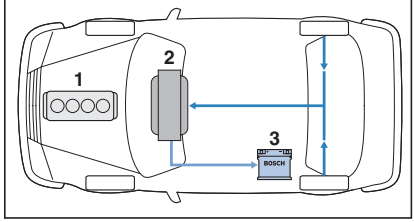


Table 1: Classification of hybrid drives

		Hybrid drive			
		Autarkic hybrid			Externally chargeable hybrid
		Start/stop system	Mild hybrid	Full hybrid	
Function	Start/stop function	●	●	●	●
	Recuperation	●	●	●	●
	Hybrid driving		●	●	●
	Purely electric driving			●	●
	Charging from the mains-supply network				●

tion together. The assistance provided by the electrical machine makes it possible to operate the internal-combustion engine by means of operating-point optimization predominantly in the range of its best efficiency or in operating ranges in which only low pollutant emissions occur. Hybrid driving can be subdivided into generator and motor modes of the electrical machine.

Generator mode
In generator mode the internal-combustion engine is operated in such a way that it delivers a higher power output than is needed for the desired propulsion of the vehicle (Figure 4). The excess mechanical power is supplied to the electrical machine and converted into electrical energy, which is used to charge the battery.

Motor mode
In motor mode, also called boosting, the electrical machine supports the internal-combustion engine in providing the desired drive power. The battery is discharged in this operating state (Figure 5).

Two different objectives are pursued with boosting: Vehicle performance can be increased compared with a purely engine-powered drivetrain. The second possibility is to maintain the existing overall drive power and use a smaller, more efficient internal-combustion engine (power-neutral downsizing).

Purely electric driving
In the case of purely electric driving the vehicle is driven solely via the electrical machine. For this operating mode the internal-combustion engine is disengaged from the vehicle drive and switched off. With the electric drive the vehicle drives quietly, without local emissions and fuel-efficiently on the energy from the battery (Figure 6).

Figure 5: Hybrid driving, motor mode
The arrows indicate the energy flow.
1 IC engine,
2 Electrical machine,
3 Battery.

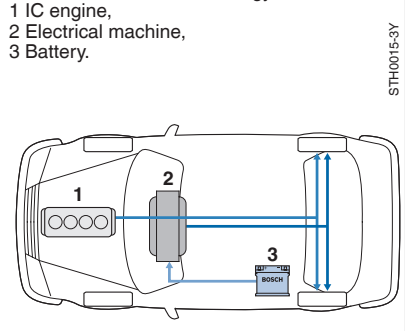


Figure 4: Hybrid driving, generator mode
The arrows indicate the energy flow.
1 IC engine,
2 Electrical machine,
3 Battery.

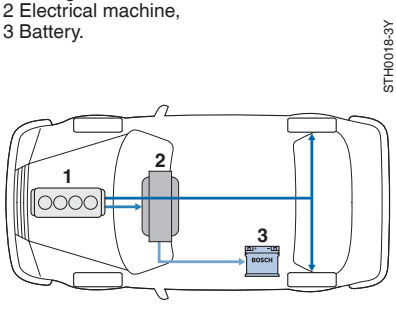
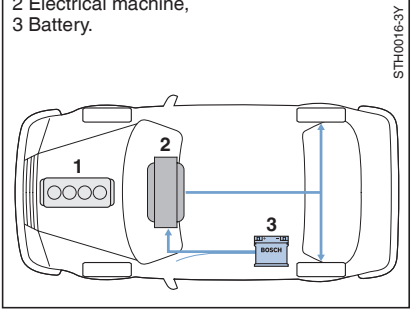


Figure 6: Purely electric driving
The arrows indicate the energy flow.
1 IC engine,
2 Electrical machine,
3 Battery.



Classification of hybrid drives

Depending on the operating state and required drive power, the internal-combustion engine and the electric drive contribute to propelling the vehicle to different extents. Hybrid drives can be classified according to the amount of power the electric drive contributes to the overall power of the vehicle and the possible functionalities (see Table 1). Vehicles can be classified according to the following features:
– start/stop system (microhybrid),
– mild hybrid,
– full hybrid,
– externally chargeable hybrid.

Start/stop system

Features

A start/stop system, also called a micro-hybrid, essentially equates to a conventional drive with internal-combustion engine, alternator as electrical machine, starter, and starter battery. The power of the electrical machine is at approximately 3 kW not high enough to make a contribution to propulsion. Therefore it is not in the proper sense a hybrid drive.

This system implements the start/stop (see Start/stop functionality) and recuperation functions. During recuperation the alternator increases at increased regulator voltage its power in the vehicle's overrun mode (with zero fuel consumption) to convert as much kinetic energy as possible into electrical energy and store it in the starter battery (12 V). During vehicle accelerations and at high load the alternator decreases its power depending on the battery charge state as far as possible by reducing the regulator voltage. In this way the load on the internal-combustion engine caused by the alternator and the fuel consumption are reduced. The alternator supplies less electrical energy in this phase; the consumers draw the electrical energy from the battery.

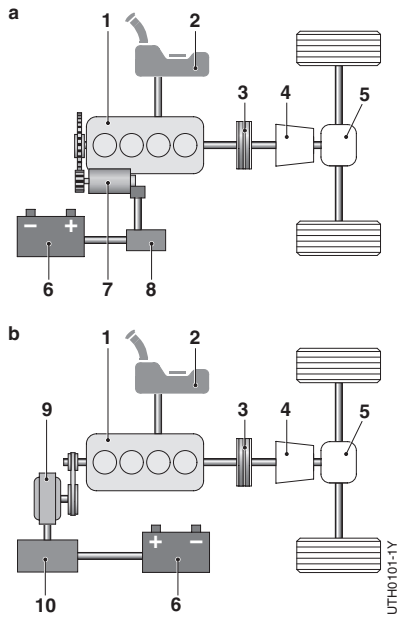
This start/stop system requires battery management with a charge-state sensor system (battery sensor). The amount of energy that can be recuperated is however limited in the 12 V vehicle electrical system.

Classic starter

The start/stop functionality can be achieved with a classic starter (Figure 7a). The engine-management system switches the engine off when the vehicle comes to a stop. The engine restarts automatically when the brake is released (automatic transmissions) or the clutch is operated (manually shifted transmissions). Starters which are designed for a higher number of starting operations are used for this application. The battery must also be designed for this operation, exhibiting increased cycle stability.

This system offers a simple way of reducing fuel consumption with limited outlay. These system with start/stop functionality and recuperation can therefore now be found in many vehicle models.

Figure 7: Start/stop system
a) Conventional system with starter,
b) System with belt-driven starter generator.
1 IC engine,
2 Fuel tank, 3 Clutch, 4 Transmission,
5 Final-drive unit, 6 Battery,
7 Start/stop-capable starter,
8 Starter control,
9 12 V starter generator,
10 Power electronics.



Starter generator
Starter generators combine the starter and the alternator (generator) in a single component. It is permanently connected to the internal-combustion engine for the generator function. When the engine starts this electrical machine drives the internal-combustion engine. The starter generator can, depending on the installed power, also support (boost) the internal-combustion engine.

Belt-driven starter generator
The belt-driven starter generator is connected via a belt drive to the internal-combustion engine (Figure 7b). The layout is the same as that of a conventional alternator; this topology can also be found in a mild hybrid with 48 V vehicle electrical system. It can be easily integrated, with no changes to the drivetrain needed. Because, depending on the operating state, there is a slack strand or taut strand in the belt drive, mechanical adjustments are however required in the belt-drive tensioning system. In the case of a belt-driven starter generator the power transmittable by the belt determines the upper limit for the system power.

Starter generators which are connected via belts to the internal-combustion engine were used in series production for the first time in 2004.

Integrated starter generator
The integrated starter generator, also known as the crankshaft starter generator, is seated between the internal-combustion engine and the transmission on the crankshaft. In the case of an integrated starter generator the transmittable power is limited by the starter generator's performance and its power electronics.

The coaxial mounting of the starter generator on the crankshaft necessitates a lengthening of the drivetrain. Thus, integrating such a system in an existing vehicle concept entails an outlay on mechanical adjustments.

The crankshaft starter generator can demonstrate its advantages particularly with voltages of 42 V and upwards. A 42 V vehicle electrical system had been discussed for some time, but was not then introduced. Hence the crankshaft starter generator could not make the breakthrough in the 12 V vehicle electrical system in parallel to the belt-driven starter generator. This concept is first implemented with the mild hybrids and the 48 V vehicle electrical system.

Advantages of a starter generator over a classic starter
Starting operation
A classic starter accelerates the internal-combustion engine to around 250 rpm. At this speed fuel is injected in such a way that the engine automatically reaches idle speed. This operation, because of the starter pinion's meshing process and its rotation in the engine's flywheel, is accompanied by noises and vibrations. Because the starter generator is permanently connected to the crankshaft, these side-effects occur much less frequently. The starter generator can also accelerate to higher rotational speeds; the first injection can thus take place at higher rotational speeds (high-speed start). This results in a quieter and faster starting operation.

Start/stop system
Because of the meshing process, a classic starter can only initiate the starting operation while the engine is stopped. This reduces the effectiveness of a start/stop system because the engine can only be restarted at zero speed. Therefore, with these start/stop systems, the engine can only be switched off when the vehicle is stationary. With a starter generator, on the other hand, the engine can already be switched off while it is coasting, which thanks to the longer engine-standstill phases impacts positively on fuel consumption. A renewed starting operation can if necessary be initiated again even when the engine is turning.

Mild hybrid

Features

With a mild-hybrid system the hybrid-driving function can be implemented in addition to the functions of a start/stop system. An electrical machine assists the internal-combustion engine here. As well as the 12 V battery, the mild hybrid contains a traction battery with a higher voltage level (e.g., 48 V) from which the electrical machine draws its energy. Purely electric driving is not possible because the electrical machine and the internal-combustion engine are permanently connected. These systems are therefore called mild hybrid (mild hybridization).

Integration in the drivetrain

There are different ways of connecting the electrical machine in the drivetrain and to the engine. Figure 8 shows the connection of the electrical machine to the internal-combustion engine via a belt drive. The electrical machine can be used in this hybrid vehicle instead of the alternator from the 12 V vehicle electrical system.

The other widespread integration option is to arrange the electrical machine on the crankshaft between the internal-combustion engine and the clutch or transmission. This arrangement facilitates the transmission-ratio-free connection of the electrical machine to the internal-combustion engine. However, the additional component increases the length of the drivetrain.

Depending on the hybrid-vehicle power requirement, there are further integration options for one or if necessary more electrical machines (see Drive structures).

Full hybrid

Features

A full-hybrid system can, in addition to the functions of a mild-hybrid system, drive over shorter distances with the electrical machine alone. The internal-combustion engine is disengaged during purely electric driving from the drivetrain by the power take-up element or an interrupting clutch and switched off.

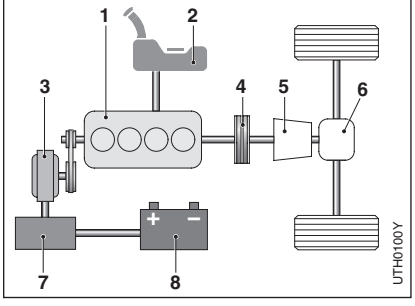
The traction battery can store energy during recuperation and with excess combustion-engine power in the case of operating-point shifts. The voltage level of the traction batteries in a full hybrid are typically up to 400 V and for high power requirements up to 800 V. However, the capacities of the traction batteries used are not sufficient to facilitate purely electric driving over longer distances. But the electric drive in a full hybrid can increase the efficiency of the internal-combustion engine.

Because the battery in the case considered here is not rechargeable from an external source, full-hybrid vehicles are from the electric-storage perspective “autarkic”. The full hybrid is therefore assigned to the group of autarkic hybrid systems (see Table 1).

The Toyota Prius was the first full-hybrid vehicle to be a volume-production model in 1997. Since then further vehicle models have been brought onto the market as full hybrids.

Figure 8: Installation position of belt-driven electrical machine

1 IC engine, 2 Fuel tank,
3 Electrical machine, 4 Clutch,
5 Transmission, 6 Final-drive unit,
7 Pulse-controlled inverter, 8 Battery.



Integration in the drivetrain
A full-hybrid vehicle uses both an internal-combustion engine and at least one electrical machine as the means of propulsion. There are in this respect a multitude of drive structures which partly pursue different optimization objectives and which utilize to differing extents electrical energy to drive the vehicle (see Drive structures).

Electrical machine
Permanent-field synchronous machines (PSM), asynchronous machines (ASM), and electrically excited synchronous machines (ESM) are used for full hybrids (see Electrical machines for motor-vehicle propulsion, types of electrical machines).

Traction battery
Battery voltages of typically 200 to 400 V are required for use in full hybrids with the given power requirements relating to the electrical machine. The system voltage can be increased by a step-up converter to up to 800 V. The battery voltage for hybrid applications in commercial vehicles, in plug-in hybrids and in pure electric vehicles can be up to 800 V. The storage capacity of the traction battery depends on the requested range that can be covered by purely electric means. Typical storage capacities of the traction batteries used in full hybrids are up to 2 kWh. Thus, purely electric driving is possible over shorter distances (typically around 2 to 3 km) with full-hybrid vehicles.

Vehicle electrical systems
Full hybrids have an HV (high-voltage) vehicle electrical system in line with the battery voltage. The low-voltage vehicle electrical system also required in full hybrids (12 V in the passenger-car and 24 V in the commercial-vehicle sector) with a battery is connected via a DC/DC converter to the HV vehicle electrical system. The battery in the low-voltage electrical system is charged via the DC/DC converter; the alternator in the conventional vehicle electrical system is not required here.

Inverter
The inverter generates from the HV battery voltage the three-phase alternating current (see Inverter) for operating the electrical machine.

Advantages
The combination of electric and combustion-engine drives in the full hybrid has the following advantages over a conventional drivetrain:
– The electric drive offers constantly high torques at low rotational speeds. In this way, it ideally supplements the internal-combustion engine, whose torque only starts to increase at mid-range rotational speeds.
– The electric drive and internal-combustion engine together are thus able to deliver a high dynamic response from every driving situation.
– The assistance provided by the electric drive makes it possible to operate the internal-combustion engine with a higher-geared transmission with identical mileages predominantly in the range of its best efficiency or in ranges in which only low pollutant emissions occur. Operating-point optimization occurs.
– The combination with an electric drive often facilitates the use of a smaller internal-combustion engine while retaining the same overall power output (power-neutral downsizing).
– Through generator operation of the electrical machine, it is possible when braking as compared to a mild hybrid to convert a much greater part of the vehicle's kinetic energy into electrical energy (recuperation). The electrical energy is stored in the energy storage device and can be used at a later stage to drive the vehicle.
– The electric drive can be use for purely electric driving. Here the internal-combustion engine is switched off and the vehicle is locally operated with zero emissions.

Externally chargeable hybrids
Plug-in hybrid
Features
An extension of the full-hybrid system is the plug-in hybrid, whose traction battery can be charged not only internally but also from the mains-supply network – i.e., externally. As a result, the journey can be started with a fully charged traction battery.

Electrical machine
The electrical machine in a plug-in hybrid is the same as the in a full hybrid.

Traction battery
The traction battery for plug-in hybrids as a rule has a larger capacity (e.g., 16 kWh) than that for full hybrids. This enables plug-in hybrids to be driven longer distances of typically up to 70 km (increasing trend) without the internal-combustion engine cut in. The decisive factor here is that a plug-in hybrid with a fully charged traction battery can be started and driven purely electrically over predominantly short distances.

Range extender
Features
The range of an electrical vehicle on one battery charge is currently still much lower than that of a vehicle with an internal-combustion engine. An added complication is that the charging process is very time-consuming, making it difficult to use an electric vehicle for longer distances. Range extenders denote additional assemblies in an electric vehicle which extend the range limited by the restricted battery capacity. Internal-combustion engines with limited power for example are used for this purpose. These vehicles are designed as electric vehicles; the power of the electrical machine and the capacity of the traction battery correspond to those of an electric vehicle. The traction battery is charged from the public mains-supply network, e.g., at a charging post. The range extender enables the vehicle to continue to be driven with a flat battery to the next charging facility.

Classification based on drive structure

The different hybrid drives can be classified on the basis of how the electrical machine is integrated in the drivetrain. The requirements which are imposed on an electrical machine depend not least on the installation location and consequently on the drivetrain topology. There are different ways of configuring the internal-combustion engine, transmission and electrical machines in hybrid vehicles. The different drive structures can be divided using possible energy flows into three categories:
– parallel hybrid drive,
– series hybrid drive,
– power-split hybrid drive.

Parallel hybrid
Topologies
In a parallel hybrid drive both the internal-combustion engine and the electrical machine influence the vehicle drive simultaneously (parallel). The electrical machine of the parallel hybrid drive can basically be positioned at different points in the drivetrain, which brings with it specific advantages and disadvantages. A nomenclature originally defined by Daimler AG has gained acceptance for designation purposes; this system designates the parallel-hybrid drivetrain based on the position of the electrical machine in the drivetrain as P0 through P4. The P here stands for the parallel architecture, while the number denotes the installation location of the electrical machine in the drivetrain.

Figure 9 shows the layout of the electrical machine for the different topologies using the example of a standard drive. An external charging socket can be installed but is not pictured here. This also applies to the front-transverse drive.

Essentially, the electrical machine can be installed in the drivetrain in a number of different positions:

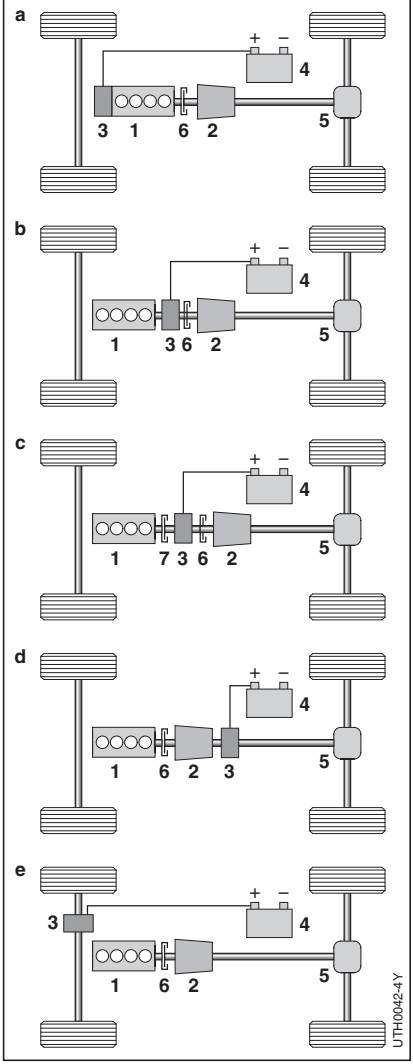
– on the internal-combustion engine with belt connection to the crankshaft (P0 topology),
– between internal-combustion engine and clutch or power take-up element, directly connected to the internal-combustion engine (P1 topology),
– between internal-combustion engine and transmission input, can be disconnected via a clutch from the internal-combustion engine (P2 topology),
– on the transmission input (P3 topology),
– connected to the final-drive unit or integrated in the wheel (P4 topology).

Features of the parallel hybrid
In parallel hybrid drives the internal-combustion engine and one or more electrical machines independently of each other or jointly contribute to propelling the vehicle (full hybrid) or to boosting the internal-combustion engine (mild hybrid). The two energy flows from the internal-combustion engine and the battery therefore run parallel to each other, the two individual components of power adding up to a total drive power. Parallel hybrid drives come in a mild-hybrid variant (with start/stop functionality, regenerative braking, and hybrid driving) or in a full-hybrid variant (additionally with electric driving).

Advantages of the parallel hybrid
A fundamental advantage of the parallel hybrid is the possibility of maintaining the conventional drivetrain in wide ranges. For the hybrid versions of the respective vehicle no new transmission has to be developed. A modular construction kit can be used to realize both conventional transmissions (dual-clutch or automatic planetary transmissions, but also continuously variable transmissions) in conjunction with internal-combustion engine and transmissions specially developed for hybrid drives with integrated electrical machines with different power stages.

The extent of development and installation work for parallel drive structures is lower when compared with series and power-split structures, in that for the most part only one electrical machine with lower electric power is required and the adaptations necessary in converting a conventional drivetrain are fewer.

Figure 9: Schematic diagram of a parallel hybrid architecture with various arrangements of the electrical machine in a standard drive
a) P0, battery and electrical machine mostly 48 V,
b) P1, c) P2, d) P3, e) P4 (axle hybrid).
1 IC engine, 2 Manual transmission,
3 Electrical machine, 4 Traction battery,
5 Final-drive differential,
6 Power take-up element (clutch or converter),
7 Interrupting clutch.



P0 topology
The electrical machine is permanently connected via a belt drive to the internal-combustion engine (see Figure 8). It replaces in this topology the conventional alternator (see also Start/stop system).

The electrical machine can with this topology perform the conventional alternator function even when the vehicle is at a standstill and thereby reliably supply the vehicle electrical system. On the proviso that the electrical machine has a suitably compact design and no additional water cooling is required, mounting the electrical machine in the original alternator position entails only minimal additional integration outlay, since the alternator installation space can be utilized. However, the belt tensioner must be adapted to the different operating states in which taut strand or slack strand occurs.

Because the electrical machine is connected to the internal-combustion engine, the engaged internal-combustion engine reduces the energy to be recuperated when decelerating.

The power output of the electrical machine connected via a belt drive to the internal-combustion engine is typically approximately 10 kW. Consequently, only a mild hybrid can be realized with the P0 topology. The CO_2 savings potential is thus also limited due to the electrical machine's restricted power output.

At an operating voltage < 60 V there is no need for the costly measures to comply with the contact-protection stipulations.

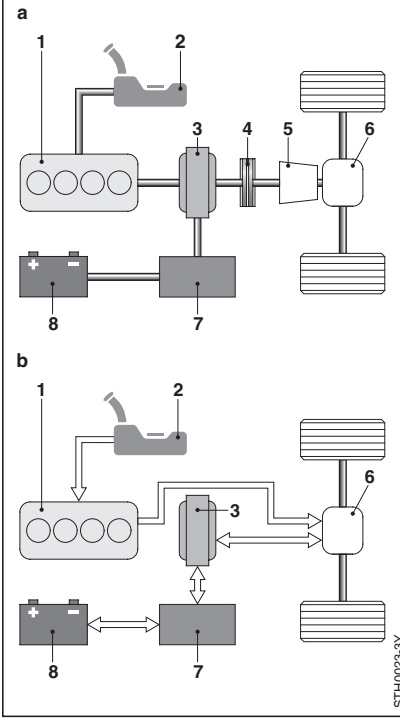
P1 topology
In the P1 topology (Figure 10) the electrical machine is again connected directly to the internal-combustion engine. The electrical machine can be connected via a belt as in the P0 topology, making it possible to realize the functions of a mild hybrid.

When the electrical machine is seated, as shown in Figure 10, directly on the crankshaft (typically a permanent-field synchronous machine), the loss-affected belt drive is omitted, consequently increasing the mechanical efficiency. Furthermore, when compared with a belt-driven electrical machine, a more powerful electrical machine can be used because the power transmittable via the belt is lower.

In contrast to series and power-split drive structures, the speed of the internal-combustion engine is permanently coupled to the speed of the electrical machine. In vehicle deceleration phases the internal-combustion engine cannot be decoupled from the electrical machine and is therefore always engaged. In the process the drag torque of the internal-combustion engine reduces the energy to be recuperated during deceleration. The CO_2 savings potential is thereby limited.

Purely electric driving is not possible with this drive structure. The electric drive can be used to assist the internal-combustion engine and thereby improve dy-

Figure 10: P1-topology parallel hybrid with one clutch
a) Drive structure,
b) Energy flow.
1 IC engine, 2 Fuel tank,
3 Electrical machine, 4 Clutch,
5 Transmission, 6 Final-drive unit,
7 Pulse-controlled inverter, 8 Battery.



namic performance. The P1 topology is therefore only suitable for a mild hybrid.

The power output of the electrical machine connected via a belt drive to the internal-combustion engine is typically approximately 10 kW. Values of around 20 kW are typical for an electrical machine integrated on the crankshaft.

The outlay on integrating the electrical machine in the drivetrain is higher here than for the P0 topology with belt drive. The P1 topology is possible with all transmission variants.

P2 topology

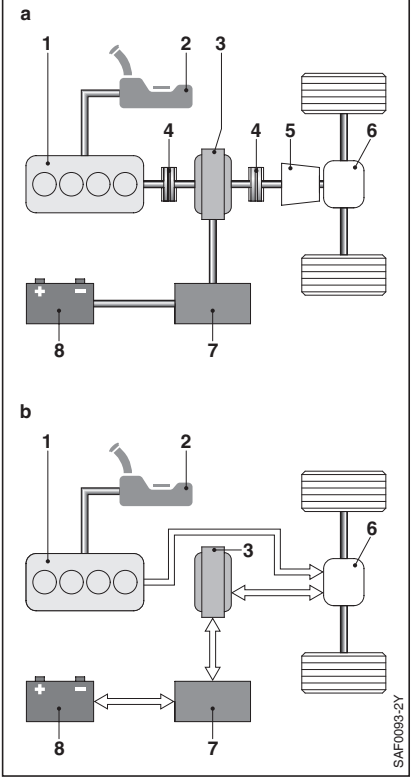
If a clutch is installed between the internal-combustion engine and the electrical machine which allows the engine to be cut in and out and thus decoupled from the electrical machine (Figure 11), purely electric driving with a full hybrid is made possible. The internal-combustion engine can also be decoupled in deceleration phases. Firstly, this increases the energy to be recuperated during acceleration. Secondly, it permits sailing operation, whereby the vehicle rolls freely and is slowed down only by aerodynamic drag and rolling friction. However, engine starting can only take place with the clutch closed, which hampers the transition from electric driving to driving with the internal-combustion engine, above all without a clutch between the electrical machine and the transmission.

To make this drive structure acceptable, it is very important to enable the internal-combustion engine to be started from electric driving without compromising comfort. There are two different ways of achieving this. In the first option the internal-combustion engine is started when the interrupting clutch is open by a separate starter and there is no unwanted effect on vehicle motion. This however calls for a separate starter, which can actually be dispensed with in a hybrid vehicle. Another option is to activate the internal-combustion engine, the electric drive and the clutch in such a way that the effect on vehicle motion is compensated for during engine starting. To this end an intelligent control system requires access to measured values from the internal-combustion engine, the electric drive, and the clutch located in between. The clutch must be able to adapt automatically to the changing conditions in ongoing operation and to follow the control-system inputs.

The additional moment of inertia acting via the transmission input shaft must be taken into account in the design of the synchronizer rings. Reliable generator operation via the internal-combustion engine is interrupted when the clutch is open.

The integration outlay for this topology is high since significant changes to the

Figure 11: P2-topology parallel hybrid with two clutches
a) Drive structure,
b) Energy flow.
1 IC engine, 2 Fuel tank,
3 Electrical machine,
4 Clutch or converter (optional),
5 Transmission, 6 Final-drive unit,
7 Pulse-controlled inverter, 8 Battery.



transmission housing are required. Installation space is additionally required for the electrical machine and the clutch. The P2 topology is possible with all transmission variants.

P3 topology
Connecting the electrical machine to the transmission outlet is also feasible. This topology is possible with all transmission types.

However, in this topology the rotational-speed spread of the drive axle must be transmitted to the machine shaft. Electric driving can only be achieved by choosing a suitable conversion ratio between the electrical machine and the drive shaft. If, at higher driving speeds, this results in the upper rotational-speed limit of the electrical machine being exceeded. the electrical machine must be automatic disengaged.

In principle, this topology offers the greatest efficiency during recuperation since the lowest mechanical transmission losses occur here. But this only applies provided the electrical-machine power output does not decrease substantially with the speed. Otherwise the maximum possible recuperation or boost power would drop prematurely with the driving speed. If the electrical machine has be to disengaged due to its rotational-speed limit at higher driving speeds, energy cannot be recuperated in these high rotational-speed ranges.

The internal-combustion engine cannot be started from a vehicle standstill. As well as the additionally required conventional alternator for reliable electrical-system supply this topology must have a conventional start/stop-capable starter or a 12 V- or 48 V-based starter generator.

The integration outlay for the electrical machine must here too be gaged as very high. Furthermore, installing the additional clutch between the internal-combustion engine and the electrical machine in this P3 topology entails increasing the length of the drivetrain and thus necessitating a larger installation space.

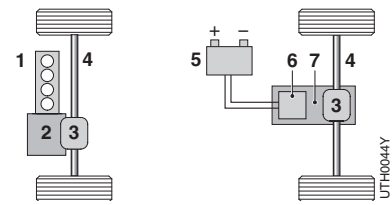
P4 topology
A further parallel drive structure is obtained by electrifying a separate axle (axle-split parallel hybrid, Figure 9e for standard drive, Figure 12 for front-transverse drive). Here, a conventional drivetrain comprising internal-combustion engine and transmission on a powered axle is combined with an electrically powered axle. A semi-automatic transmission and a start/stop system for the internal-combustion engine are required for this purpose. One or more interrupting clutches can be installed.

This drive structure is classed as a parallel hybrid drive because the individual components of power from the engine and the electric drive add up. This concept delivers a four-wheel-drive functionality limited by the power of the electrical machine and battery charge. Distribution of the drive torques can be adjusted by specific activation of the electric drive within wide limits.

The traction battery is recharged by regenerative braking or externally via a charging socket. The traction battery can only be recharged externally when the vehicle is stationary.

The transmission for the internal-combustion engine is the same as that of a vehicle without axle hybrid, but must also be able to be operated over longer distances at high speed in neutral – i.e., open. Lu-

Figure 12: Schematic diagram of an axle hybrid (P4) in a front-transverse drive
An external charging socket can be installed but is not pictured here.
1 IC engine, 2 Manual transmission,
3 Final-drive differential, 4 Drive shaft,
5 Traction battery, 6 Electrical machine,
7 Step-up gear unit (fixed ratio).



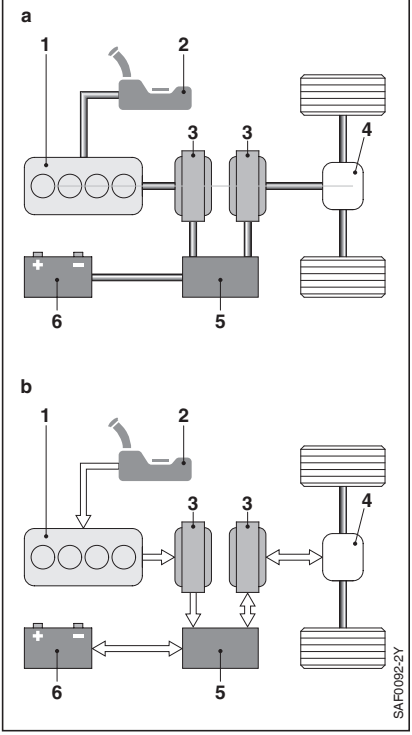
brication of the transmission must be assured when both the input shaft and the output shaft are stationary.

Final drive via an axle
Particularly for front-transverse drives with very limited installation space it can be advantageous not to have to accommodate the electrical machine between the engine and the transmission. A P0 or P1 position by means of a belt drive is just as conceivable as a P4 position on the axle driven by the internal-combustion engine. The downside of this concept is that the electrical machine may not be able to utilize the different gears of the combustion-engine transmission and is subject to losses at high driving speeds as long as it cannot be disengaged from the final-drive differential and the wheels.

Series hybrid
In a series-hybrid drive there is no mechanical connection of the internal-combustion engine to the vehicle drive. The internal-combustion engine drives an electrical machine, which operates as an alternator (Figure 13). The electric power generated in this way is made available together with the battery power to a second electrical machine, which drives the vehicle. From the perspective of the energy flows, a series connection exists in this case.

Advantages of the series hybrid
In view of the fact that there is no mechanical connection between the internal-combustion engine and the powered wheels

Figure 13: Series hybrid drive
a) Drive structure,
b) Energy flow.
1 IC engine, 2 Fuel tank,
3 Electrical machines, 4 Final-drive unit,
5 Pulse-controlled inverter, 6 Battery.



in the series hybrid, this drive structure offers some advantages. Thus, a conventional range transmission is not required in the drivetrain. This creates free space for arranging the overall drive ("packaging").

In addition, starting the internal-combustion engine from electric driving does not cause any unwanted effect on vehicle motion.

The chief advantage when driving is the ability to freely select the operating point of the engine. Excess mechanical energy produced by the internal-combustion engine at full load with high efficiency which is not required for vehicle propulsion can be stored in the battery and then utilized later when the combustion engine is turned off for the electric drive. This supports fuel-saving and low-emission operational management of the vehicle. The engine can also be optimized to a limited operating range.

With a larger battery capacity the electrical machine operated as an alternator and the internal-combustion engine can be designed with smaller dimensions. The combustion engine can be maintained at more favorable operating points; it only makes available the average energy required for vehicle propulsion for the base load. The missing power for acceleration processes for example is drawn from the battery.

As the combustion engine and alternator further decrease in size, this system devolves into a range extender.

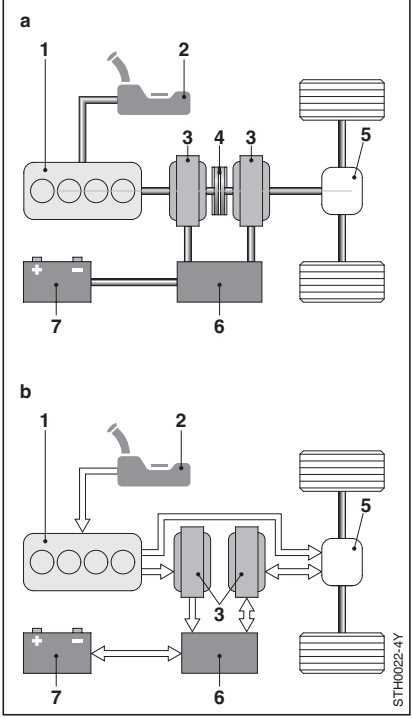
Disadvantages of the series hybrid

A downside of the series hybrid is that electrical energy has to be converted twice. The losses incurred through double energy conversion are higher than in the case of a purely mechanical transfer by a transmission. Furthermore, this principle lacks the possibility of adding the power of the internal-combustion engine to that of the electrical machine (boosting).

Standing against the high driving comfort provided by the electrical machine without ratio change and the possibility of operating the internal-combustion engine at its optimum consumption point is the high complexity in the form of two electrical machines. These, due to their very concept, have to be very powerful since the entire energy of the internal-combustion engine must be converted. This means greater weight and higher costs.

Due to the double energy conversion, the series hybrid is only fuel-efficient under certain operating conditions. At low and medium speeds a series hybrid, despite the higher losses, offers a fuel-consumption advantage because here the positives afforded by the ability to freely select the engine's operating point outweigh the negatives. The higher losses are predominant at higher speeds.

Figure 14: Series-parallel hybrid drive
a) Drive structure,
b) Energy flow.
1 IC engine, 2 Fuel tank,
3 Electrical machine, 4 Clutch,
5 Final-drive unit, 6 Pulse-controlled inverter,
7 Battery.



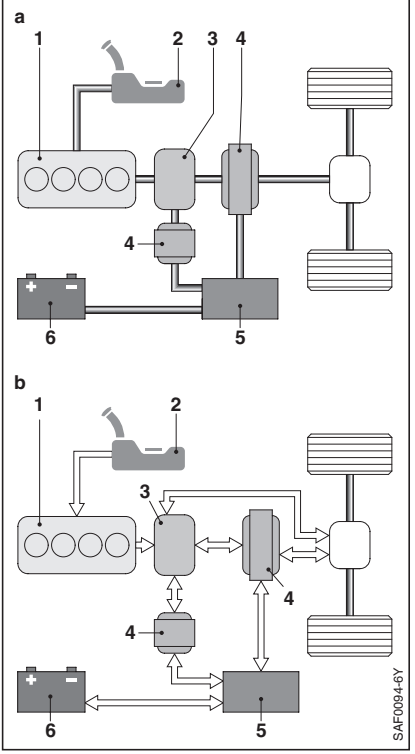
Series-parallel hybrid
A series hybrid is extended to become a series-parallel hybrid (Figure 14) by establishing a mechanical connection between the two electrical machines which is connected or separated by a clutch. The series-parallel hybrid can utilize the advantages of the series hybrid at low speeds and circumvent the disadvantages at higher speeds by closing the clutch. When the clutch is closed, the series-parallel hybrid behaves like a parallel hybrid. Since the double energy conversion is limited to the range of lower speeds and power outputs, smaller electrical machines are sufficient for the series-parallel hybrid than are required for the series hybrid.

Compared with the series hybrid, the advantage in "packaging" is lessened on account of the mechanical connection between the internal-combustion engine and the powered wheels. Compared with the parallel hybrid, two electrical machines are required for the same task.

Power-split hybrid
Features
A power-split hybrid drive generally has two electrical machines, which on the one hand function as electric variators and on the other hand perform "hybrid" functions such as boosting and recuperation. Power-split hybrid vehicles combine features of parallel and series hybrid vehicles with those of a power split (Figure 15). For the variator function of the two electrical machines some of the mechanical power (speed times torque) is supplied via the planetary gear to the first electrical machine, which now functions as a generator and converts the mechanical power into electric power (electrical path). The remaining power, together with the second electrical machine, drives the vehicle (mechanical path). The electric power is supplied via an inverter to the second electrical machine, which functions as a motor at a speed independent of both the generator and the internal-combustion engine and supplies its power to the drivetrain again in mechanical form.

Power split
The central element of power split (or power summation) is a planetary-gear set, which is connected via its three shafts to the internal-combustion engine and two electrical machines. For the input-power split the internal-combustion engine is connected to an element of the planetary-gear set (e.g., the planetary-gear carrier), the generator to a second element (e.g., the sun gear), and the second electrical machine to the third element (e.g., the ring gear) and the output. For the output-power split the internal-combustion engine and the generator engage an element of the planetary-gear set and

Figure 15: Power-split hybrid drive
a) Drive structure,
b) Energy flow.
1 IC engine, 2 Fuel tank,
3 Planetary gear, 4 Electrical machine,
5 Inverter, 6 Battery.



the second electrical machine engages a second element of the planetary-gear set, while the output is connected to the third element of the planetary-gear set. In both cases the electric component in the power-split hybrid can be varied by activation of the electrical machine. This provides for continuously variable starting and continuously variable ratio adjustment. The battery can be charged or discharged by varying the generated and supplied electrical energy of the two electrical machines.

Because of the kinematic boundary conditions at the planetary-gear set, the engine speed can be adjusted within certain limits independently of the driving speed. The internal-combustion engine can thus operate extensively in the optimal operating range. Following a continuously variable transmission (CVT), the terminology used here is an electric continuously variable transmission (ECVT).

Some of the engine power is transmitted by the planetary-gear set via the mechanical path to the powered wheels. The remainder of the power is transmitted via the electrical path with double energy conversion to the powered wheels. Similarly to the series hybrid, the electric transmission path can be used in the case of low power demands. For higher power demands, the mechanical transmission path is also available. However, it is not possible to switch arbitrarily between the mechanical and electrical transmission paths. Depending on the configuration of the planetary-gear set, the electrical machine and the internal-combustion engine, without an additional transmission only certain combinations between the mechanical and electrical transmission paths are ever possible. In this way, the power-split hybrid can achieve substantial fuel savings at low and medium speeds. No additional fuel savings can be achieved at high speeds.

The power-split hybrid can be expand by mechanical, fixed gears by using a second planetary-gear set. The mechanical complexity increases, while the electrical complexity decreases. Smaller electrical machines are then sufficient for a comparable concept. Furthermore, fuel consumption at medium and higher speeds can be improved.

Disadvantages

Standing against the high driving comfort of this concept is the high complexity in the form of two electrical machines, which due to their very concept have to be very powerful as they must convert a large proportion of the power. Their overall power is thus in the range of the installed combustion-engine power. There is no provision for modularity with an existing basic transmission without electrical machines.

Incorporating the transmission in the hybrid drive

Transmissions for P0 and P1 topologies

Hybrid drives come in the most varied forms. For P0 and P1 topologies, where the electrical machine is located between the internal-combustion engine and the clutch or converter, the classic drivetrain components can be used. The required changes in the drivetrain for the hybrid drive are minimal; no changes to the transmission itself are required. All transmission types can be used.

Adapted conventional transmissions

Conventional transmissions adapted to hybrid functionality are prevalent in vehicles with P2-topology hybrid drive. These vehicles are based on the variant of classic drive with internal-combustion engine; the classic drive components can be used for the most part. The function of the automatic transmission is the same as that of a conventional drive.

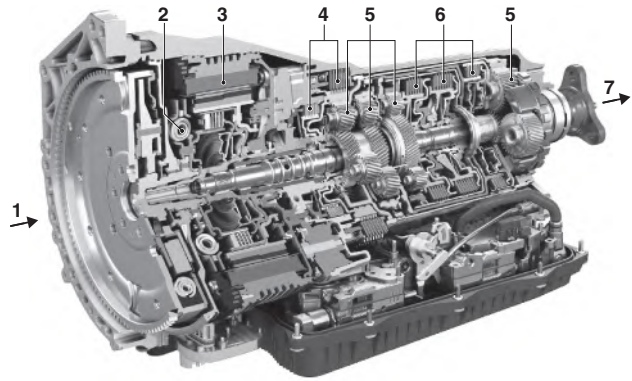
Figure 16 shows an automatic planetary transmission for the standard drive in whose housing an electrical machine is integrated at the drive end (IMG, integrate motor-generator, see Electrical machines for motor-vehicle propulsion). With these transmissions it is possible to achieve a modular, gradual electrification of the drivetrain ranging from the mild hybrid, through the full hybrid up to the plug-in hybrid. This transmission with integrated electrical machine is only slightly longer than a conventional transmission. The extended length even fits into the same outer transmission case on frequent occasions. This simplifies the vehicle platform.

Transmissions with integrated electrical machine

In some vehicles the installation space required for the P3 drive configuration is not available. This problem can be remedied by integrating the electrical machine in a dual-clutch transmission. The electrical machine is no longer connected to the crankshaft of the internal-combustion en-

Figure 16: Structural design of a P2 transmission with integrated electrical machine base on a conventional automatic planetary transmission for a standard drive

1 From driving engine, 2 Torsion damper, 3 Electrical machine, 4 Multi-disk brakes, 5 Planetary-gear sets, 6 Multi-disk clutches, 7 To final-drive unit.



gine, but is instead connected to a subunit of the dual-clutch transmission. This arrangement does not require an additional clutch between the internal-combustion engine and the electrical machine. Purely electric driving with the engine stopped is possible by opening the double clutch of the transmission. Depending on the gear selected in the transmission subunit, with the electrical machine a different transmission ratio is possible between the engine and the electrical machine. This produces an additional degree of freedom for the hybrid control, which can be used to further reduce fuel consumption.

Special transmissions for hybrid drives

Specially developed hybrid vehicles which are only offered as hybrid variants arrive on the market in high numbers. Here it is practical to adapt the entire drivetrain to the hybrid-drive possibilities and thereby optimize the complete system. One measure is to simplify the mechanical design of the transmission by omitting the components for the reverse gear. Instead, at least one electrical machine is integrated into the transmission to provide for reversing (backing up). The electrical machine is thus part of the transmission, whereby the connection to the transmission can be effected by various transmission shafts. The transmission can be further simplified by reducing the gears. Tight stepping of the gears as encountered in a classic drive is not necessary in a hybrid drive.

Transmissions specifically developed for a hybrid drive with one or more integrated electrical machines called dedicated hybrid transmissions (DHTs) [3]. They can be derived from all the possible transmission concepts – the dual-clutch transmission, the automatic planetary transmission, the continuously variable transmission (CVT) or the automated manual transmission. Power-split operating conditions can be created with planetary-gear sets additionally integrated in such transmissions.

References

[1] Directive 2007/46/EC of the European Parliament and of the Council of 5 September 2007 establishing a framework for the approval of motor vehicles and their trailers, and of systems, components and separate technical units for such vehicles (Framework Directive).

[2] M. Uhl: 3rd International Conference Automotive 48 V Power Supply Systems. 24 – 26 November 2015, Düsseldorf.

[3] H. Tschöke, P. Gutzmer, T. Pfund (Editors): Die Elektrifizierung des Antriebsstrangs. 1st Edition eBook, Verlag Springer Vieweg, 2019.

Control of a hybrid drive

Hybrid vehicle

In a hybrid electric vehicle (HEV) propulsion is provided by an internal-combustion engine and one or more electrical machines (e-machines). Depending on the performance of the electrical machine there is mild hybridization with a mild hybrid (battery voltage typically 48 V) or strong hybridization with a full hybrid (also called a strong hybrid, battery voltage typically 200 to 400 V; HV High Voltage). The following components are required for the drive of a hybrid vehicle:

– internal-combustion engine,
– clutch to separate the electrical machine from the internal-combustion engine,
– transmission (e.g., automatic planetary transmission, dual-clutch transmission),
– electrical machine,
– power electronics (inverter),
– battery (e.g., 48V battery for mild hybrid, HV battery for full hybrid),
– DC/DC inverter for connecting the HV vehicle electrical system to the low-voltage electrical system,
– regenerative braking system for recuperative braking.

Figure 1 shows by way of example a vehicle with electric final-drive unit (e-axle).

Vehicle control unit for hybrid vehicles

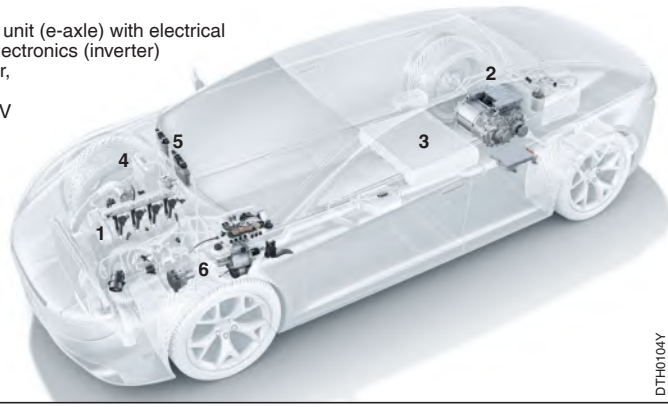
Function

In a hybrid vehicle these components must be operated by open- and closed-loop control. The efficiency which can be achieved with the relevant hybrid drive is crucially dependent on the higher-level hybrid control. Figure 2 shows the electrical system overview of a hybrid drive with the networking of the individual components and ECUs in the drivetrain. The higher-level hybrid control (vehicle control unit, VCU) coordinates the entire system, the subsystems of which have their own control functions. These are battery management, engine management for the internal-combustion engine, management of the electric drive (including the electrical machine, the clutches, and the power electronics), transmission management, and management of the braking system.

In addition to pure control of the subsystems, the hybrid control also includes an operating strategy which optimizes the way in which the drivetrain is operated. The operating strategy brings influence to bear on the consumption- and emission-reducing functions of the hybrid vehicle; it must implement the different hybrid functions, such as regenerative braking and hybrid and electric driving, and start-

Figure 1: Components in the drivetrain of a hybrid drive with electric final-drive unit

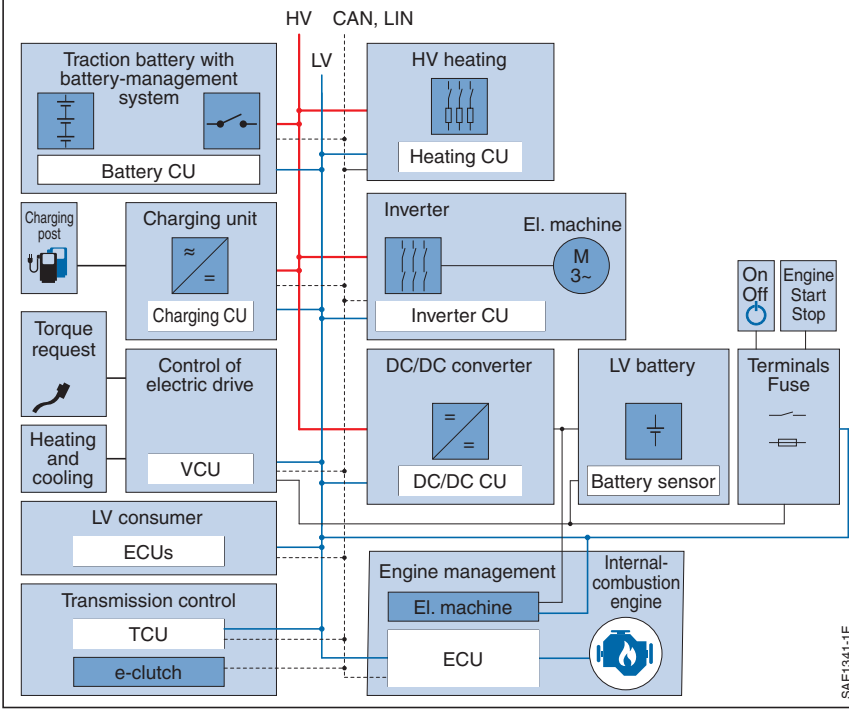
1 IC engine,
2 Electric final-drive unit (e-axle) with electrical machine, power electronics (inverter) and reduction gear,
3 Traction battery,
4 12V battery for 12V vehicle electrical system,
5 Control units,
6 Regenerative braking system.



stop operation of the internal-combustion engine. The operating strategy determines how the drive power is shared between the engine and the electric drive. It decides the extent to which the potentials for fuel saving and for reducing the emissions of a vehicle are utilized.

Selection of and switching between the individual states is performed in accordance with numerous conditions relating to, for example, the accelerator-pedal position, the state of charge of the battery, and the vehicle's current speed. The components in the hybrid vehicle behave differently, depending on the optimization objective (e.g., fuel saving or emission reduction).

Figure 2: Electrical system diagram for a hybrid vehicle
HV High Voltage,
LV Low Voltage (12 V for passenger cars, 24 V for commercial vehicles),
AC Alternating Current,
DC Direct Current,
CAN Controller Area Network, LIN Local Interconnect Network,
CU Electronic Control Unit,
VCU Vehicle Control Unit (electric-drive control unit),
TCU Transmission Control Unit,
ECU Engine Control Unit.



Drivetrain operating strategy for hybrid vehicles

Function

In a hybrid vehicle with internal-combustion engine and electrical machine the interaction of these two drives must be coordinated. This drivetrain operating strategy is part of the software in the VCU (vehicle control unit), which controls the degrees of freedom of the interaction of the drives in such a way as to minimize the overall drive-system losses over the route.

The overall wheel torque requested (or the wheel force or the mechanical power) is determined by a driver request (accelerator-pedal position, vehicle-speed controller). The purpose of the optimization strategy is to delivery this overall torque with the best possible distribution between the internal-combustion engine and one or more electrical machines. The strategy must furthermore decide whether it is more favorable to drive by hybrid or purely electric means if purely electric driving is possible for the given hybrid-vehicle topology.

Classification of operating strategy

Generally, the operating-strategy solutions can be classified into three main categories: control-based, optimization-based, and learning strategies.

Control-based operating strategy

Control-based operating strategies comprise deterministic strategies and fuzzy-logic strategies. In the deterministic strategy events – particularly future events – are clearly defined by preconditions. In typical fuzzy-logic strategies the continuous data are classified into discrete value ranges. Both strategies have a series of predefined rules without the remaining route being known from the outset.

Optimization-based operating strategy

Optimization-based strategies can be either offline or online approaches. In the case of offline approaches optimization is performed offline; the results are then stored in the VCU for example in program maps. Dynamic programming (DP) is an established offline approach. This involves splitting the optimization problem into subproblems with the intermediate results being stored. The optimal solution of the overall problem is made up of the optimal solutions of the subproblems.

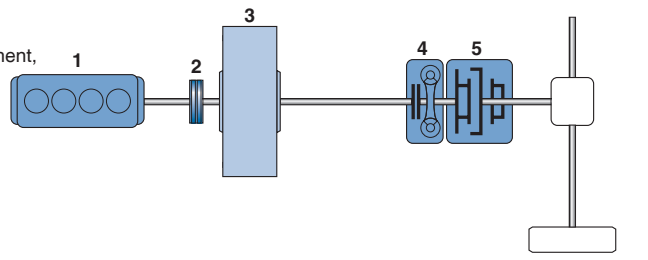
In the case of online optimization strategies, on the other hand, an algorithm for calculating the optimal operating points is calculated directly in the VCU. Examples of online optimization approaches are the equivalent consumption minimization strategy (ECMS) andmodel predictive control (MPC) [1]. MPC approaches predict a trajectory of vehicle motion, calculate a cost function for each step in the trajectory, and select the nearest element ahead of the vehicle for present optimization [2]. Online approaches can react better to different driving conditions.

Learning strategies

Learning strategies demonstrate that promising advances are as a rule possible with approaches based on neural networks [3].

Figure 3: Topology of a P2 parallel hybrid

1 IC engine,
2 Interrupting clutch,
3 Electrical machine,
4 Power take-up element,
5 Transmission.



Influences on operating strategy

The description of a drivetrain operating strategy applies here generally to all hybrid topologies. The concept of an operating strategy is explained with reference to a P2 parallel hybrid (Figure 3). The P2 topology features a interrupting clutch between the internal-combustion engine (ICE) and the electrical machine, which in turn is connected via a transmission to the wheels. The arrangement of the drives facilitates both purely electric and hybrid driving. In the hybrid driving mode the operating strategy determines the optimal torque distribution or the optimal power distribution. The operating strategy also decides between operation in the hybrid driving mode and the electric driving mode whereby the total losses during optimal power distribution in the hybrid driving mode are compared with the total losses in the electric driving mode. It is important to bear in mind that the operating strategy must find an optimum for each instant taking account of the overall distance traveled and not just for an instantaneous optimization. This is required because the battery state of charge (SOC) is a long-term aspect of the system. Discharge of the battery resulting from electric driving for example can only be evaluated if the system takes into account how much electrical energy is returned to the battery (e.g., by load-point increase, operation of the internal-combustion engine with excess power for generator operation of the electrical machine).

Figures 4 and 5 show representative speed/torque program maps of an internal-combustion engine and an electrical machine. It becomes clear that electrical machines operate efficiently in a wide speed and torque range, i.e., with high efficiency. The internal-combustion engine on the other hand is much less efficient and has only certain speed/torque ranges with higher efficiency which are marked in Figure 4 as the optimal operating curve. Values with the same specific fuel consumption (values in g/kWh) form the characteristic curves shown in the diagram. The program map is also called a graph diagram. Optimization of the operating point of internal-combustion engine (Figure 4), electrical machine (Figure 5), and battery and taking into account the relevance of the instantaneous driver-command torque in relation to the energy requirement for the entire journey give rise to a complex optimization problem.

Figure 4: Example of the efficiency of a gasoline internal-combustion engine depending on speed and torque

The optimal torque versus speed is marked as the optimal operating curve [4].
1 Optimal operating curve,
2 Values for specific fuel consumption in g/kWh.

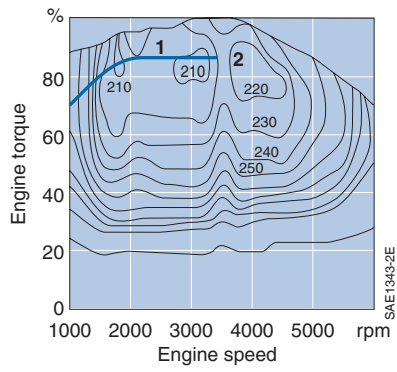
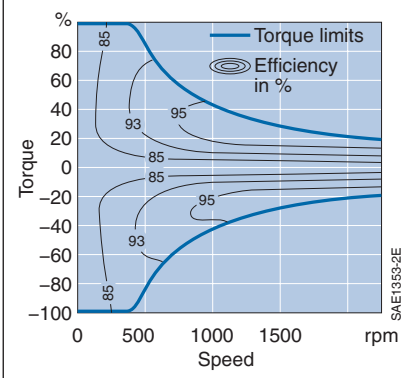


Figure 5: Example of the efficiency of an electrical machine depending on speed and torque within their physical limits [5]

Torque > 0: motor mode,
Torque < 0: generator mode.



Equivalent consumption minimization strategy

A method for solving the previously explained problem is an equivalent consumption minimization strategy (ECMS). The term "equivalent consumption" comes about as a result of the ECMS attempting to make consumption of electrical energy from the battery comparable with fuel consumption. This strategy optimizes the torque distribution or the power distribution between an electrical machine and an internal-combustion engine by attempting to store and utilize electrical energy in the battery with the highest possible global efficiency (overall efficiency of all the system elements involved). The ECMS achieves this by taking account of losses of the internal-combustion engine, the electrical machine (including inverter), and the battery.

The idea behind the ECMS is to find a way to compare the power delivered by the internal-combustion engine with the power of the electric system. This is achieved by using an equivalence factor s comprising several aspects for optimizing the torque distribution between engines. The equivalence factor will be described in detail later.

The power requested by the driver P_D is the sum total of the mechanical power of the internal-combustion engine P_{ICE} and that of the electrical machine P_{EM}.

$$P_D = P_{ICE} + P_{EM} . \quad (1)$$

The ECMS is based on a cost function (equation (2)) in which the equivalence powers P_{Eq} are considered as costs. This cost function must then be minimized:

$$P_{Eq} = P_F + s\, P_B . \quad (2)$$

P_F and P_B are the power figures that the system must deliver to achieve the desired power [6], s is the equivalence factor. P_F is the power obtained from the fuel for the internal-combustion engine; it is obtained from the sum total of the engine power P_{ICE} and the engine losses P_{ICEL}. P_B is the power drawn from the battery; it is obtained from the sum total of the mechanical power of the electrical machine P_{EM} and the losses in the machine P_{EL}.

Consequently:

$$P_{Eq} = P_{ICE} + P_{ICEL} + s\,(P_{EM} + P_{EL}) . \quad (3)$$

The desired optimum from equation (3) is the point at which the overall system losses are minimized.

The losses P_{ICEL} and P_{EL} are generally determined empirically by way of program maps depending on the speed and the torque of the internal-combustion engine and the electrical machine, as shown in Figures 4 and 5. The values of P_{ICE} and P_{EM} for the operating points at which the losses are minimized are determined by an optimization algorithm which evaluates a multitude of P_{ICE} and P_{EM} combinations and stores the combination for which P_{Eq} assumes the lowest value.

Figure 6 shows some results by way of example of an ECMS optimization. The minimum equivalence power constitutes here the optimal torque distribution based on the cost function. Figure 6a shows the equivalence power for the hybrid drive in kW (y-axis) plotted against the torque of the electrical machine in Nm (x-axis) at an internal-combustion-engine and electrical-machine speed of 1500 rpm and a total torque demand of 120 Nm. The shift towards a negative electrical-machine torque (i.e., generator operation and charging of the battery) with a lower SOC becomes clear in Figure 6a. The equivalence power for electric driving at which the driver's overall torque request is effected via the electrical machine and the internal-combustion engine is isolated and switched off is also shown in Figure 6a for two SOC values. At the higher SOC value for this combination of driver-requested torque and drivetrain speed the equivalence power for electric driving is lower than the best possible torque distribution for hybrid driving.

Figure 6b shows for two SOC values the equivalence-power levels for hybrid drive and electrical-machine torque at an internal-combustion-engine and electrical-machine torque of 3000 rpm and a total torque demand of 400 Nm.

Equivalence factor

The equivalence factor s plays a key role in the ECMS by providing a comparison

between the power from the battery and the fuel. The equivalence factor is typically subdivided into two parts: a constant s_0 and an SOC-based correction. The s_0 component depends on the future speed and the load profile of the vehicle. Because this information can only be estimated, a correction of s_0 is made based on the deviation between the desired and the actual SOC of the battery. This also serves to maintain the SOC in a range so that charging or discharging is always possible where required. Figure 7 shows a potential realization of the behavior [3].

One drawback of the ECMS is that to calculate the optimal power distribution in the hybrid driving mode equation (3) generally cannot be mathematically solved for the minimum. The equivalence power must therefore be calculated several times in order to find the optimal solution. This results in a high computing-capacity requirement [7].

Selection of the optimal operating mode
As already mentioned, the strategy must select which operating mode – hybrid or purely electric driving – is optimal. To find the best solution, the equivalence power with the best power distribution in the hybrid driving mode is compared with the equivalence power in the electric driving mode. The prerequisite for this is that the driver request can be completely fulfilled by the electrical machine and the internal-combustion engine has no losses because it is isolated and switched off. The mode in which the lowest equivalent overall power can be achieved is then selected as the optimal energy mode.

Figure 7: Possible behavior curve of the dependence of equivalence factor s as a function of state of charge (SOC)
1 Example of the SOC end value for a driving cycle.

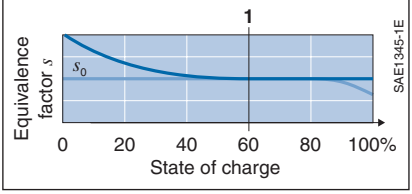
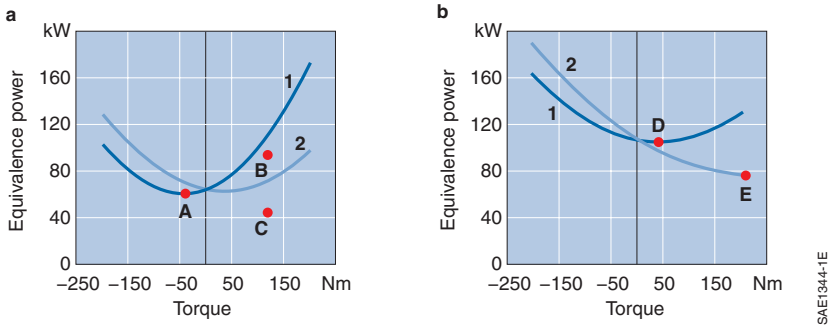


Figure 6: Cost function (equivalence power in kW vs. torque of electrical machine in Nm) for some exemplary operating conditions for a P2 hybrid vehicle (source: Bosch)
a) Transition from hybrid to electric driving, speed 1500 rpm, overall driver command 120 Nm,
b) Transition from electric to hybrid driving, speed 3000 rpm, overall driver command 400 Nm.
1 Hybrid driving, state of charge SOC = 30 %,
2 Hybrid driving, state of charge SOC = 80 %.
A Cost minimum for hybrid driving with SOC = 30 %,
B Costs of electric driving with SOC = 30 %,
C Costs of electric driving with SOC = 80 %,
D Cost minimum for hybrid driving with SOC = 30 %,
E Cost minimum for hybrid driving with SOC = 80 %.



Map-based strategies
A common type of control-based strategy is a map-based strategy. In this approach, the optimal torque distribution for example from program maps can be determined on the basis of driving conditions. The program maps which are to be selected to determine the torque distribution are determined on the basis of a set of rules which are dependent on system properties such as for example SOC classification (high, medium, low).

Map-based strategies offer the following advantages and disadvantages compared with an optimization-based operating strategy, for example an ECMS approach.

Advantages of a map-based strategy compared with the ECMS
– Freedom from calibration offers the implementer when calibrating the data (e.g., program maps) more possibilities to determine the behavior of the drivetrain in certain driving situations (e.g., combination of SOC, driver request, gear selection).
– Less computing time in real time is required because the decision as to which program maps are to be used and calling up the desired value from the program maps take place only once per optimization time step.
– In the solution it is less probable that the optimal torque distribution will exhibit great leaps with minimal changes to the driving conditions.

Disadvantages of a map-based strategy compared with the ECMS
– Major demands are made of data storage, some series solutions have more than 100 calibration curves and maps to cover all the combinations of relevant input parameters adequately.
– It is not possible to transfer the calibration from one vehicle model to another because the calibration of the individual curves and maps is based not on the physical properties of the individual components but on the combined effects of all the components in the system. In comparison, an ECMS calibration typically consist primarily of program maps that show the loss characteristics of the internal-combustion engine, the electrical machine and the battery individually.
– There is no capacity to react to changes in the environmental conditions. Depending on the implementation an ECMS can adapt online if the equivalence factor changes on account of a long-term deviation between the desired and the actual battery SOC.

On the basis of these advantages and disadvantages both equivalent consumption minimization strategies (ECMS, i.e., optimization-based strategies) and map-based strategies are used by a number of vehicle manufacturers.

Outlook – Artificial intelligence and neural networks
Neural networks (NNs) simulate the workings of the brain, permit a kind of biological information processing, and are a technical concept for solving different types of problems, including pattern recognition, prediction, planning, etc. [8]. Neural networks are the usual modules of artificial intelligence. A neural network consists of a combination of
– neurons which are interconnected (Figure 8),
– weights which define the interactions between inputs to the neural network and the neurons and interactions between neurons.

A neural network is adapted by a training process in which a lot of data are used with combinations of inputs to the neural network and resulting “correct” outputs. The training process adapts the weights in the neural network to be able to depict all the training data in the best possible way (see for example [8]).

Furthermore, progress is being made in the field of neural networks and their use in recognizing situations and trends and in solving optimization with appropriate accuracy. Generally speaking, data-based processes like neural networks have the reputation of imposing high demands on computing capacity and data storage. Interestingly, studies have shown that an efficiently designed neural network can deliver virtually optimal results for an

operating strategy for hybrid drives with fewer online computing resources and data-storage requirements than an ECMS method [3].

General design of a neural network for a hybrid operating strategy

As already mentioned, the P2 topology facilitates two driving modes, namely hybrid driving and purely electric driving. Therefore the solution of the neural network must not only determined the optimal mechanical power of the electrical machine in the hybrid driving mode but also provide information to facilitate the decision as to whether the electric or hybrid driving mode is better from the point of view of energy optimization. A neural network with three outputs for example could be selected to solve this optimization problem (Figure 8):

– The mechanical power of the electrical machine equating to the best solution in the hybrid driving mode (P_{H_O}),
– the equivalent overall power with the best power or torque distribution in the hybrid driving mode (P_{H_E}), and
– the equivalent overall power in the electric driving mode assuming that the driver's torque request is completely fulfilled by the electrical machine (P_{E_E}).

The decision as to which driving mode (hybrid or electric driving) is to be selected can then be made in the software outside the neural network, where for example the energy required for the transition of the driving mode and for driver-acceptance reasons the permissible time between the transitions are taken into consideration (the driver is informed of the mode change typically via a display).

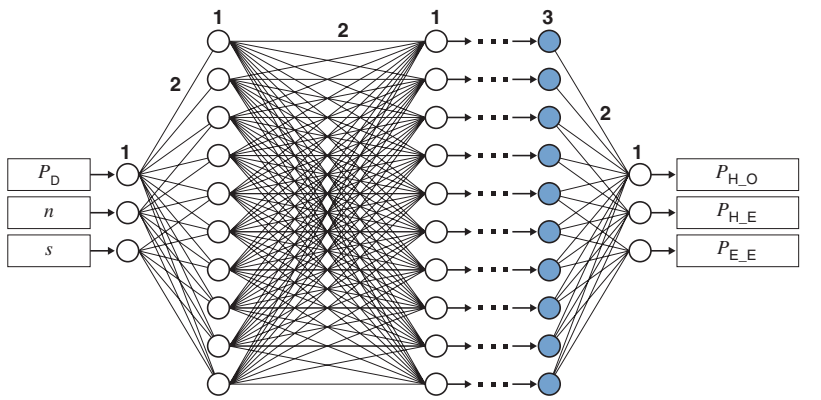
As shown in Figure 8, the network inputs are the power requested by the driver (P_D), the speed n of the internal-combustion engine and the electrical machine, and the equivalence factor s.

Application

To obtain data for the training, data records, i.e., combinations of values for P_D, n and s, were generated on the basis of a general test plan (design of experiments, DOE), with which the entire solution space

Figure 8: Concept of a neural network for a drivetrain operating strategy of a hybrid drive

1 Neurons,
2 Connections (synapses),
3 Neurons for expansion to a plastic neural network.
P_D Mechanical power requested by the driver,
n Speed of internal-combustion engine and electrical machine,
s Equivalence factor,
P_{H_O} Optimal mechanical power of the electrical machine in the hybrid driving mode,
P_{H_E} Equivalent overall power in the hybrid driving mode with optimal power distribution,
P_{E_E} Equivalent overall power with optimization of the costs of purely electric driving.



is depicted by combinations of P_D, n and s. The resulting data records were then fed into an ECMS model which was already calibrated for the target vehicle in order to supply the "correct" answers for training the neural network. In addition, test cases were created for the neural network using international driving cycles in order to represent real driving situations.

Use of plasticity in an NN-based HEV operating strategy

Networks with the ability to remember and to learn from experienced have garnered a lot of attention in the last few years. A promising form of network was published in [9]. The authors introduced a neural network with plastic weights, i.e., weights that follow Hebb's rule from biology and mean that the connections between neurons (synapses) change depending on the intensity of their interaction [10]. By applying plasticity to artificial networks it is possible to learn even after the first training and thereby facilitate adaptation to actual situations. Compared with LSTM networks (Long Short-Time Memory) the architecture of plastic neural networks is much less complex and requires less memory capacity, which makes them ideal for use in an electronic control unit for a vehicle.

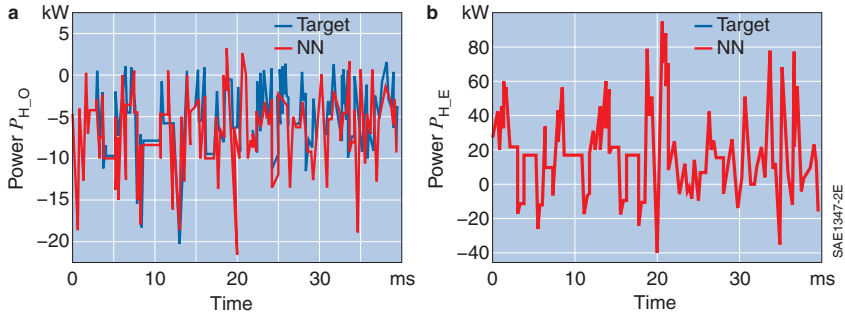
The shaded neurons in Figure 8 are affected by the expansion to a plastic neural network [3].

The results are shown in Figure 9. Figure 9a shows the direct comparison of the resulting optimal mechanical power of the electrical machine over time. Figure 9b shows the results when the mechanical NN power is converted into an equivalent power and compared with the equivalence power from the basic ECMS result. There is no discernible difference here between the NN result and the ECMS. Figure 9b shows that – although the NN results for the optimal mechanical power of the electrical machine deviate from those of the ECMS result – the deviation with regard to energy consumption makes no difference. This is useful because under certain conditions a large difference in the power or torque distribution can only have a very small influence on the overall energy consumption.

The neural network was able to deliver results comparable with an ECMS approach with regard to global efficiency while is demonstrated much fewer computation steps than the ECMS and much fewer data-storage requirements than a map-based strategy. The addition of the plasticity terms also reduced the overall system error while the advantages with regard to storage capacity and computing time were retained [3].

Figure 9: Comparison of the optimal mechanical power of the electrical machine in the hybrid driving mode from ECMS (target data) and from the NN approach

a) Optimal mechanical power of the electrical machine,
b) Comparison of the resulting optimal equivalence power for mechanical power of the electrical machine as the result of the neural network versus optimal equivalence power from the ECMS (the curves differ so little that they overlap).



References

[1] D.D. Tran, M. Vafaeipour, M. El Baghdadi, R. Barrero, J. Van Mierlo, O. Hegazy: Thorough state-of-the-art analysis of electric and hybrid vehicle powertrains – Topologies and integrated energy management strategies, Renewable and Sustainable Energy Reviews, 119, 2020, 109596 (2020).

[2] Y. Huang, H. Wang, A. Khajepoura, H. Heb, J. Ji: Model predictive control power management strategies for HEVs – A review. Journal of Power Sources, Vol. 341, pp. 91-106 (2017).

[3] K. Hauer, J. Girard: Energy Management of Hybrid Electric Vehicles Based on Plastic Neural Networks. Submitted for publication, International Journal of Powertrains (2020).

[4] K. Chau, Y. Wong: Overview of power management in hybrid electric vehicles. Energy conversion and management, Vol. 43, number 15, pp.1953–1968, Elsevier

[5] X. Hu, N. Murgovski, L. Johannesson, B. Egardt: Energy efficiency analysis of a series plug-in hybrid electric bus with different energy management strategies and battery sizes. Applied Energy, Vol. 111, pp. 1001–1009, Elsevier (2013).

[6] A. Sciarretta, M. Back, L. Guzzella: Optimal control of parallel hybrid electric vehicles. IEEE Transactions on control systems technology, Vol. 12, No. 3, pp. 352–363, 2004, IEEE.

[7] A. Panday, H.O. Bansal: A review of optimal energy management strategies for hybrid electric vehicle. International Journal of Vehicular Technology, Vol. 2014, Hindawi.

[8] O. Frölich: Einführung in Neuronale Netze. Wintersemester 2005/2006 VO 181.138 Einführung in die Artificial Intelligence, Technische Universität Wien, (2005).
https://www.dbai.tuwien.ac.at/education/AIKonzepte/Folien/NeuronaleNetze.pdf

[9] T. Miconi, J. Clune, K. Stanley: Differentiable plasticity: training plastic neural networks with backpropagation. Proceedings of the 35th International Conference on Machine Learning (ICML2018), Stockholm, Sweden, PMLR 80, 2018, arXiv:1804.02464 [cs.NE].

[10] D. O. Hebb: The organization of behavior – a neuropsycholocigal theory. A Wiley Book in Clinical Psychology., pp. 62–78, John Wiley & Sons Inc., 1949.

Electric final-drive unit

The electric final-drive unit (e-axle, known by Bosch as eAxle) is a compact electric drive solution for battery-electric vehicles and for fuel-cell and hybrid vehicles. The three relevant components of an electric drive – the power electronics (inverter), the electrical machine (motor), and the transmission – are combined in a functional unit (Figure 1).

Thanks to the functional integration of power electronics, electrical machine and transmission, the complexity can be significantly reduced in contrast to three separate components.

Functional component integration

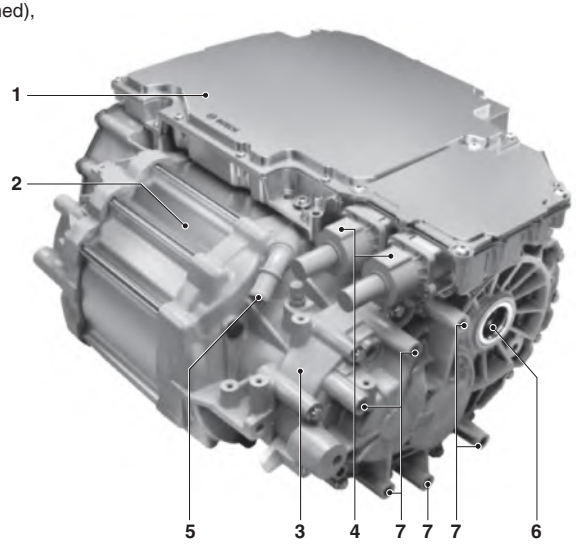
Functions of components

The power electronics, also known as an inverter, is required to generate from the direct voltage of the vehicle's traction battery via the resulting direct current (DC) a frequency-variable alternating voltage or a 3-phase alternating current (AC) for the electrical machine. A rotating field and with it a torque is generated in the electrical machine with the aid of this 3-phase alternating current.

The mounted transmission serves to transmit the torque for the vehicle drive. A two-stage spur-gear drive is usually used. The transmission has a fixed step-down ratio between 8 and 13 and a differential for compensating different wheel speeds, e.g., when cornering. Multi-speed solutions may also be used as an alternative to the solution with a fixed step-down ratio. Continuously variable transmissions (CVTs) are also possible.

Figure 1: Design of the electric final-drive unit (e-axle) from Bosch

1 Power electronics (inverter, horizontally aligned),
2 Electrical machine,
3 Transmission,
4 HV-connection to traction battery (HV: High Voltage),
5 Connection to cooling circuit,
6 Differential output (connection of side shafts),
7 Mechanical connecting points.



Manual transmissions with up to four gears are also used in the commercial-vehicle sector with the aim of providing a high starting torque for the acceleration process and an operating point with optimum efficiency at constant speed.

By converting the electrical energy, the power electronics also regulates the entire electric drive to the required setpoint torque and monitors the components for safe operation, e.g., by measuring the critical component temperatures and further system limit values.

The design and the function of the power electronics and the electrical machine are described in detail elsewhere (see Inverters, electrical machines for propulsion).

Construction

The electrical machine and the transmission are permanently joined to each other at the housing end via a common end plate (end-face housing component). To transmit the torque the electrical machine's rotor shaft is connected with the transmission input shaft, usually via a spline.

Alternatively, the rotor and the transmission can also be connected with each other via a common uninterrupted shaft, this arrangement often being used in smaller performance classes (< 100 kW) or oil-cooled electrical machines.

The end face pointing towards the transmission is also called the A-bearing side and the face pointing away from the transmission the B-bearing side.

The power electronics is integrated in the unit comprising electrical machine and transmission. The electrical connection between the power electronics and the electrical machine is established at the end pointing away from the transmission (B-bearing side) via three conductor bars (in the case of a 3-phase electrical machine). The conductor bars have flexible elements that guarantee the compensation of tolerances during installation and operation. The conductor bars are usually connected via screwed-crimped or welded joints, where for reasons of easy exchangeability and maintenance of the components detachable connections are preferred. The modularity in the e-axle design thereby achieved also enables the three components to be freely combined and configured, thus allowing the drive system to be individually adapted to the vehicle requirements.

Thermal layout

Thanks to the integration of the three components in a single drive system the e-axle's cooling system can be specifically matched and optimized to the system requirements. The e-axle is connected to the vehicle cooling circuit.

The coolant is as a rule routed through the e-axle serially (Figure 2). If flows through first the power electronics, then the electrical machine (in the case of a coolant-cooled electrical machine), and finally the transmission (if a heat exchanger is installed). Alternatively, several cooling circuits can be deployed in parallel, e.g., if the power electronics and the transmission are in each case connected separately to the vehicle's cooling system.

Aside from the described cooling concepts, there are also special solutions such as for example pure passive cooling via the surroundings or air specifically flowing round or through the components.

Cooling of the power electronics

The power electronics is usually cooled by the coolant used in the vehicle (e.g., glycol/water mixture). To this end the coolant is routed through a cooler (e.g., pin-fin cooler with pin-shaped heat sink) inside the power electronics, which is connected with the loss-affected components (as a rule power modules and link capacitor). Consequently the heat loss from these components is dissipated in the best possible way.

Cooling of the electrical machine

The electrical machine can be cooled by the coolant in a water jacket or directly by the transmission oil by means of heat exchanger. When coolant is used, this is directed through channels in the housing jacket (e.g., helix- or meander-shaped). This solution is therefore also known as water-jacket cooling.

As an alternative to water-jacket cooling the oil from the transmission that is used to lubricate and cool the components

there may also be used. In this case, the electrical machine and the transmission share a common oil circuit with the oil routed directly into the inside of the machine and consequently also reaching the winding. Because the oil use is electrically non-conductive, it can be routed specifically to the electrical machine's heat sources. As a result, despite having a generally poorer cooling effect and higher viscosity, it delivers the necessary heat dissipation. The oil is supplied to the rotor via the rotor shaft and to the stator via grooves and channels. In this case, the arrangement is called a wet-running electrical machine. The key advantage of a wet-running electrical machine is the cooling of the winding overhang for greater performance.

Because of the direct heat dissipation, generally higher continuous outputs can be achieved with this solution with minimal design and engineering effort. In addition, with the omission of the water jacket, the electrical-machine housing can have a simpler and more compact design. It is, however, essential to ensure the chemical resistance of the materials wetted with oil, particularly the copper conductors and the insulation materials. In addition, the oils used for cooling can as a rule only be used up to a maximum temperature of approx. 150 to 180 °C. This must be taken into account for the permissible stator temperature. Because the stator in a water-cooled machine with maximum short-time performance can also reach temperatures of over 200 °C, the short-time performance may in contrast have to be reduced in an oil-cooled machine.

Both cooling media can furthermore be used in combination, e.g., cooling the jacket with coolant and cooling the rotor with oil. But, in view of the increased sealing effort involved, this combination is not very widely used.

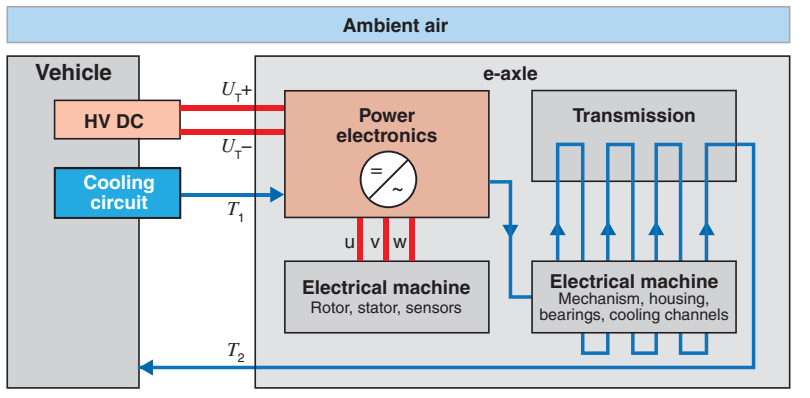
Cooling of the transmission

The oil used for lubrication in the transmission absorbs the heat loss from tooth-flank contacts. The oil must be cooled as a result. A heat exchanger, in which the heat if the oil is dissipated to the coolant, is usually used for this purpose.

Here too different cooling concepts are used to suit the performance class. In low performance classes (< 100 to 150 kW) the transmission oil is often cooled purely by passive cooling via the transmission housing with the aid of fins. In mid and high performance classes solutions integrated into the housing, e.g., channels and pin structures, can be used to dissipate the heat of the transmission oil to the

Figure 2: Thermal layout of the e-axle

T_1 Inlet temperature of coolant into the e-axle,
T_2 Outlet temperature of coolant after passing through the e-axle,
U_T+, U_T– Traction voltage (HV direct voltage, e.g., from the battery),
HV High Voltage, DC Direct Current.



coolant. Oil-to-coolant heat exchangers integrated into the oil sump are also used.

Alternative topologies
Aside from the described design, which is also referred to as axially parallel because the output shaft or the differential output is arranged parallel to the electrical machine's rotor shaft, there are further e-axle topologies.

Coaxial e-axle topology
In the coaxial e-axle topology the differential output is positioned coaxially to the electrical machine's rotor shaft. The torque at the end facing away from the transmission (B-bearing side) is provided by an additional intermediate shaft, which is guided by the hollow rotor shaft. The transmission can in this topology take the form of a two-stage spur-gear drive with countershaft or a single- or multi-stage planetary gear. The advantages of the coaxial topology are the compact design and consequently small installation-space requirement. The disadvantages are the increased costs owing to the mechanical complexity.

Arrangement of the power electronics
Aside from the axial arrangement, the topology of the e-axle is determined by the position of the power electronics. A distinction is made here between jacket-side and end-face installation.

In the case of jacket-side installation, the power electronics is positioned on the electrical machine's jacket surface. Here the power electronics can be vertically or horizontally aligned (see figure 1). Crucial to optimum alignment are, aside from the vehicle architecture, the desired accessibility in the installed state and the connection of the HV (high-voltage) conductor bars and the signal cables.

In the case of end-face installation, the power electronics is located on the end face of the electrical machine pointing away from the transmission (B-bearing side). This position known primarily from small three-phase AC drives in vehicles (e.g., radiator fans, transmission oil pumps, electric air-conditioning compressors) is also used in e-axles of smaller performance classes (< 100 kW).

T-shaped topology
Another topology that is being increasingly used is T-shaped topology. Here the transmission forms the central unit in the middle of the e-axle and accommodates the other two components. The electrical machine and the power electronics are mounted at the transmission end and the opposite end respectively. The HV conductor bars and the signal cables are routed past but very close to or even through the transmission. The advantage of this topology is that it uses side shafts of equal length without the need for an intermediate shaft. The drawbacks are the large installation-space requirement and the thermal effects of the transmission on the power electronics.

Special solutions
As well as the widely used e-axle topologies described here there are numerous special solutions, but these will not be considered any further here.

Vehicle integration and interfaces

The e-axle can be installed in the vehicle both on the front axle and on the rear axle. All-wheel applications are also possible accordingly. When it comes to the positioning of the e-axle a distinction is made between two arrangements, with the position of the electrical machine being the critical factor. This can be positioned either facing the center of the vehicle (or the traction battery) or facing away from it (Figure 3).

Arrangement of the e-axle in the vehicle

Arrangement facing the center of the vehicle

Characteristic of the arrangement facing the center of the vehicle is the position of the electrical machine between the two vehicle axles, i.e., within the wheelbase along similar lines to a mid-engine arrangement (Figure 3a).

The differential output of the e-axle usually lies in the cross-sectional plane of the vehicle axis. This position can however also be varied slightly according to the approach angles of the side shafts that can be achieved by the shaft joints, for example to create more space for the traction battery.

The key benefit of the arrangement facing the center of the vehicle is the small installation-space requirement of the e-axle in the area in front of the front axle and thus in the crash zone. In addition, further components can be arranged in front of the e-axle on the front axle, e.g., the steering box or the air-conditioning compressor. The drawback is the large installation-space requirement between the axles, which is usually used for the battery.

Arrangement facing away from the center of the vehicle

In the arrangement facing away from the center of the vehicle the electrical machine is situated in the area in front of the front axle, the upshot of which is that more space can be made available between the axles for the battery.

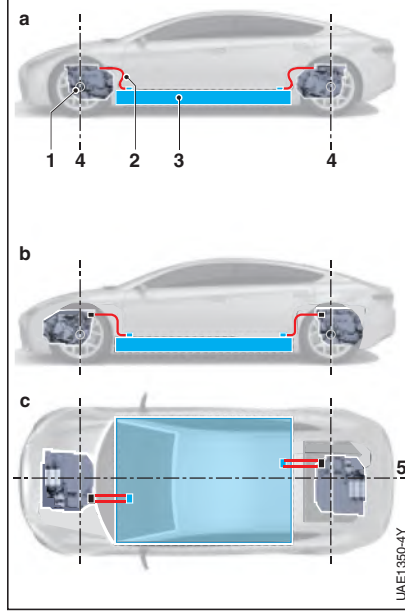
Choice of arrangement in practice

There is essentially no preferred solution when it comes to choosing the e-axle arrangement. This is very much dependent on the vehicle architecture. Both arrangements are therefore employed in virtually equal measure.

In each case the crash safety of the HV components must be taken into consideration. If the power electronics is installed upright on the jacket side, it is important to ensure that it is positioned facing the center of the vehicle and therefore does not impinge on the crash zone in the area in front of the front axle. This applies to all the arrangements described because

Figure 3: Arrangement of the e-axle in the vehicle

a) Electrical machine facing the traction battery,
b) Electrical machine facing away from the traction battery (side view),
c) Electrical machine facing away from the traction battery (top view).
1 e-axle,
2 HV cable (cable routing with S-bend),
3 Traction battery,
4 Plane of vehicle axis,
5 Center line of cross-sectional plane.



damaged cables must be exchanged completely in the event of an accident.

Interfaces
The following interfaces are featured between the e-axle and the vehicle: Electrical interfaces for high voltages (HV, >60 V) and low voltages (LV, 12 V for passenger cars, 24 V for commercial vehicles), signal / communication interfaces, and hydraulic and mechanical interfaces.

Electrical interfaces
The electrical HV connection to the traction battery is usually achieved with cables. The connection to the e-axle can be made by connectors or screw contacts. For safety reasons it is important when laying the cables to ensure that these have sufficient leeway to enable the e-axle to move particularly in the event of a crash and prevent them from ripping off and leaving open contacts. Reference is made in the cable routing to an S-bend (S-shaped cable routing) (see Figure 3).

The electrical connection to the e-axle power supply from the LV vehicle electrical system (e.g., for the electronics) and the communication interface to the vehicle (e.g., via CAN) are usually bundled in a connector with several pins. Often also featured are further vehicle-individual signal or power supplies (e.g., crash signal, HV interlock for monitoring correct connection of the plug-in connections in the HV circuit). The ground connection is established by a conducting connection from the e-axle housing to the vehicle's body structure.

Hydraulic interfaces
The hydraulic connection to the vehicle cooling system is achieved by hoses and hose clamps or by special quick-release locks.

Mechanical interfaces
The mechanical interfaces include the connection of the side shafts to transmit the torque to the wheels and the mounting of the e-axle in the vehicle body. The vehicle-side joints of the side shafts are inserted with splines into the two differential outputs. Alternatively, other joints may also be used, e.g., bolted flanges.

The e-axle is usually mounted in the vehicle body by means of bolted rubber mounts. The rubber mounts reduce the forces on the vehicle structure generated by the dynamic torque support of the e-axle and isolate the e-axle acoustically from the vehicle structure. Consequently, the structure-borne noise emitted by the e-axle can only be transmitted to a small extent to the vehicle structure.

Depending on the vehicle architecture, an auxiliary frame can also be used which contains the connection of the e-axle and the chassis. This auxiliary frame is in most cases also connected via bolted rubber mounts with the vehicle body, which brings about a double acoustic isolation of the drive from the vehicle body.

The compatibility of the e-axle with different vehicle-individual connection points is achieved through a number of bolting points on the e-axle (see Figure 1), which form by way of additional bolted brackets (support elements) the connection to the vehicle body. Furthermore, thanks to the modularity of the e-axle, the housing parts to be joined can also be fashioned individually for each vehicle and the e-axle can thus be connected directly with the vehicle body without the use of brackets.

Performance classes and performance characteristic

Performance classes

The e-axle can be individually adapted by means of scaling and the efficient combination of components to the vehicle segment and usage scenario. Based on current market requirements there is a performance spectrum of 50 to 300 kW, with the following performance classes having been established within this spectrum:

– ≤ 110 kW
– 120 to 150 kW
– 180 to 200 kW
– ≥ 250 kW

The largest shares are currently in the performance classes up to 150 kW, which are primarily used in the vehicle segments A (subcompact cars) through C (mid-size cars).

Performance characteristic

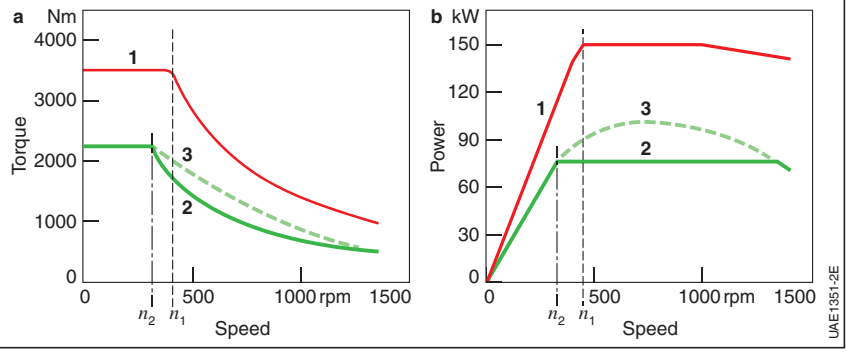
The provision of power to the e-axle is determined by the performance characteristic of the three components, where the respective limitations of the components overlap (by superposition) and in their totality define the performance characteristic of the e-axle. Added to this are possible limitations of the connecting parts, e.g., the AC conductor bars or the DC connection of the traction battery and the EMC filter components (electromagnetic compatibility).

The performance characteristic of the e-axle is derived from the design (installation space and layout) and the thermal interactions of the components used. Integrating the components in the e-axle offers advantages, but the thermal interactions must be taken into account.

In the provision of power a distinction is basically drawn between short-time performance (provision of power / maximum torque, as a rule between 10 s and 30 s operating duration) and continuous performance (usually 30 min operating duration) (Figure 4). The exact definition of short-time and continuous performance including the boundary conditions can vary from manufacturer to manufacturer.

Crucial to the short-time provision of the maximum torque particularly in the base-speed range (approx. between standstill and specific speed n_1, n_2 in Figure 4) is the provision of the maximum current by the power electronics. It must also be ensured during the short period of 10 s that the heat loss generated due to the switching losses and the ohmic losses – particularly in the power modules and in the link capacitor, but also in the other current-carrying com-

Figure 4: Performance characteristic of a Bosch e-axle (by way of example with transmission ratio i = 11.8)
a) Torque against rotational speed,
b) Output against rotational speed.
1 Short-time performance (10 seconds),
2 Real continuous performance (30 minutes),
3 Theoretical continuous performance (30 minutes).
n_1, n_2, Specific speeds.



ponents – can be dissipated and an overheating of the components is prevented. The maximum current provided by the power electronics leads in the electrical machine to the requested torque, where further component-specific boundary conditions must be observed, e.g., maximum permissible current in the electric conductors and dissipated of the heat loss arising here to the housing or the cooling medium. The maximum torque provided is crucial to the design of the transmission and other mechanical components (e.g., rotor shaft and splines).

Owing to the compact function integration in the e-axle a sufficient heat-loss dissipation within the permissible driving-function areas poses a particular challenge. In this connection, local overheating, which leads to aging and damage mechanisms within the power electronics, electrical-machine and transmission subassemblies, must be effectively prevented under different driving and operating conditions. To ensure this, various temperature sensors whose signals provide the necessary input variables for installed software component-protection models are placed inside the e-axle. These installed models help to provide comprehensive component protection inside the e-axle which depending on the different driving and operating conditions automatically ensures a thermal overload regardless of driver intervention.

The heat loss arising in the electrical machine must be dissipated in the best way possible by the cooling medium. This gives rise to the theoretical continuous performance of the electrical machine. Furthermore, the transmission's cooling concept is designed for a defined continuous output to prevent the permissible oil temperature from being exceeded. The real continuous performance of the e-axle is obtained with these boundary conditions (Figure 4). Compared with the theoretical continuous performance it is reduced above the specific speed, on the one hand for component-protection reasons and on the other hand to be able to represent a uniform output over a larger speed range. Continuous performance can be further increased by effective cooling concepts and measures, where however always the relevant application must be taken into consideration and it is necessary to weigh which continuous output is required in the respective application.

The dominant performance characteristic depends on usage. If the e-axle operates as the main drive, continuous performance is more relevant. In the event that it acts as an auxiliary drive on the axle not driven by the main drive (e.g., all-wheel drive, boost mode), short-time performance depending on the repetition rate is more relevant.

Efficiency

As well as the targeted adaptation of the performance characteristic to the respective application, another development objective is to reduce the losses of the three components and thus to increase system efficiency. To determine system efficiency, it is necessary to multiply the efficiency maps of the components by each other. Through the targeted optimization of the individual components in the system context, it is thus possible to achieve an efficiency of up to 96 % at selected operating points in the e-axle.

Furthermore, reducing the losses has a positive effect on the performance characteristic, particularly continuous performance. When the power dissipation of a component is reduced, that component is heated less with the result that with constant cooling power the limit for continuous performance is increased or cooling power can be reduced.

Scalability

Through scaling and the efficient combination of the components, the torque and output of the e-axle can be matched to the specific requirements with discrete stages in the combination of the components having been established along similar lines to the discrete performance classes listed at the outset. A certain degree of continuous scaling is possible within these stages.

The output or the maximum current made available in the power electronics is defined here primarily via the chip surface area of the power modules. The peripheral components of the power electronics (e.g., link capacitor and conductor bars) are also adapted accordingly.

The achievable electrical-machine torque can be scaled by means of the machine's diameter or length. The greater the diameter or the length, the greater the torque for the same current and same number of turns of the electrical machine. Furthermore, the torque and output of the electrical machine can varied by varying the number of turns (see Electrical machines for motor-vehicle propulsion).

It is possible by varying the step-down ration to vary between the maximum torque and the maximum rotational speed. The higher the step-down ratio, the higher (for the same electrical machine) the torque and the lower the maximum rotational speed. Varying the step-down ratio does not affect the maximum output.

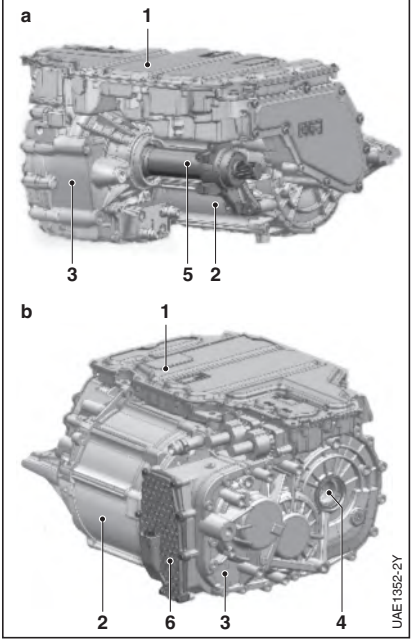
Adaptation for application

Additional optional elements can furthermore be integrated in the e-axle. These are described in brief in the following.

Intermediate shaft
Because the e-axle output shafts are not situated centrally in the vehicle in relation to the installation position, an intermediate shaft (Figure 5a) can be used to obtain side shafts of equal length and a higher number of common parts for the vehicle manufacturer. Beyond the pure cost advantage, this is also a prerequisite for symmetrical driving dynamics.

Parking lock
The parking lock (Figure 5b) is a positive mechanical lock which prevents the vehicle

Figure 5: Additional e-axle components
a) View with intermediate shaft,
b) View with parking lock.
1 Power electronics, 2 Electrical machine,
3 Transmission, 4 Differential output (connection of side shafts), 5 Intermediate shaft,
6 Parking lock (as option).



from accidentally moving (cf. position P on automatic or dual-clutch transmissions).

The parking lock is preferably actuated electrically in the e-axle. Hydraulic and pneumatic actuations are also possible, but this is only useful if the appropriate peripherals are provided, for example in the case of a two-speed solution. The parking lock can also be actuated via a mechanical cable.

The translatory and rotatory vehicle masses are mechanically locked by means of a parking-lock gear (permanently connected with a transmission shaft) and a parking-lock pawl (housing-affixed). This ensures that the driver leaves the vehicle in a safe state.

To prevent the vehicle from coming to an uncontrolled or an unwanted abrupt stop, the parking lock can no longer be engaged above a so-called ratchet or pawl speed (approx. walking speed). This is usually implemented by a mechanical rejection of the pawl at the parking-lock gear.

The customary parking brake used to date in vehicles with manually shifted transmissions generates, in contrast to the parking lock, a frictionally engaged prevention operation which cannot prevent the vehicle from rolling further in unfavorable conditions (e.g., laden vehicle with trailer on a hill). But frictionally engaged parking brakes are also used in electric vehicles.

Disconnect or disengagement unit

The electrical machine can be disconnected from the vehicle's side shafts with the aid of a disconnect unit in order to eliminate the drag losses of the electrical machine and also those of the transmission when the vehicle is rolling. This offers, in particular in the case of a permanent-magnet synchronous machine (PSM), the advantage that this does not rotate and consequently no voltage is induced and no kinetic energy of rotation is converted. This makes coasting possible, whereby the vehicle rolls without its own propulsion power.

Furthermore, it can also be beneficial in all-wheel or hybrid applications to disengage the auxiliary drive and thereby minimize the losses. During the disengagement process, the electrical machine is synchronized without torque in the disengagement unit, rotationally disconnected, and brought to a controlled standstill. During the reconnect process, the electrical machine is synchronized accordingly without load to the current rotation speed of the rotating transmission or side shaft and then connected. Positive clutches have become established here.

The disengagement unit can essentially be implemented on each of the rotating shafts inside the e-axle or also on the side shafts. In this way it is possible to disengage the rotatory masses including their friction points and thus also eliminate the transmission's drag losses.

When a PSM is used, it is also possible to tow the vehicle rolling on its wheels without inducing a voltage which otherwise is normally reduced for example by an active short circuit. As a result, the mechanical and thermal system loads are reduced and towing is made easier or made possible at all.

With multi-speed transmissions it is also possible to disconnect via the clutch provided anyway.

Connecting the air-conditioning compressor

As described already, the e-axle is isolated from the vehicle structure, as a result of which the emitted structure-borne noise is reduced. This property, together with short mechanical power-transmission paths, offers the possibility of mounting the air-conditioning compressor directly on the e-axle. The e-axle's high-voltage distributor is used for the electrical power supply. Unlike an internal-combustion engine, there is the additional benefit of lower thermal load by the electric drive because the latter produces no waste combustion heat.

Electric drives with fuel cell

Features

Fuel-cell electric vehicles (FCEVs) are electric vehicles. They enable the climate-neutral carriage of persons and goods by using regeneratively produced hydrogen. The fuel-cell electric drive (Figure 1) consists of the electric drive, as is known from the battery-electric drive, and the fuel-cell system. At the heart of the fuel-cell system is the fuel-cell stack, which consists of individual fuel cells.

Fuel cells are electrochemical converters which convert the chemical energy of the fuel directly into electrical energy. There is a large number of different fuel-cell types that use different fuels and components. For motor-vehicle applications currently the low-temperature proton-exchange membrane fuel cell (LT-PEMFC; temperature < 100 °C), which is run on hydrogen (H_2) and oxygen (O_2), is the most interesting. In such a fuel cell the hydrogen reacts with oxygen to create water, releasing electrical energy in the process.

Advantages of fuel cells

Within the field of electromobility fuel-cell drives have the advantage of permitting the accustomed convenience of conventional vehicles such as short refueling time and long range with existing H_2 infrastructure. As in an internal-combustion engine energy storage and conversion are separated, which makes increased range possible by tank expansion. In particular, fuel-cell drives therefore have an advantage when it comes to vehicles which call for high availability, useful load, and range. Examples include trucks, city and intercity buses, trains and ships, as well as commercially used vehicles such as taxis and company cars.

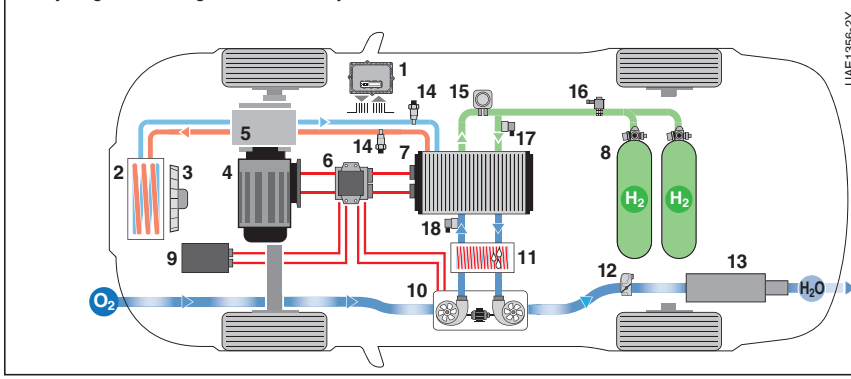
Fuel cells do not produce any greenhouse gases or air pollutants locally during operation.

Expansion of the fuel-cell system with a battery

Fuel-cell vehicles are electric vehicles in which the energy for the electric drive is made available by the fuel-cell system. For several reasons, it is sensible to expand the drivetrain to include a traction battery:

Figure 1: Drivetrain of a fuel-cell electric passenger car

1 Fuel-cell ECU, 2 Radiator, 3 Radiator fan, 4 e-machine, 5 Transmission, 6 Power electronics, 7 Fuel-cell stack, 8 Hydrogen tank, 9 HV battery, 10 Air compressor, 11 Humidifier, 12 Pressure-control valve, 13 Muffler, 14 Temperature-pressure sensor, 15 Anode recirculation fan, 16 Tank shutoff valve, 17 Hydrogen-metering valve, 18 Safety valve.



– the energy gained through recuperative braking can be temporarily stored,
– the dynamic response of the drivetrain can be further increased,
– the efficiency of the drivetrain can be further increased by shifting the load point of the fuel-cell system.

The ratio of battery power and total power varies for different applications. Typically, fuel-cell systems are used as the primary energy source in the drivetrain. Such vehicles are characterized as strong fuel-cell electric vehicles.

Alternatively, the battery can have a much greater power rating and energy content, and be recharged by a smaller fuel-cell system where necessary. This drive concept is called fuel-cell range extender because it allows the range of battery electric vehicles (BEVs) to be increased.

Fuel-cell system power ratings of up to 100 kW are usual in today's FCEVs in the passenger-car segment. Batteries have a power rating of up to 40 kW with an energy content of 1 to 2 kWh. In the commercial-vehicle sector either the higher fuel-cell power demand (200 to 400 kW) is realized by modular approaches (combination of several lower-power modules) or the fuel-cell system is directly configured for the higher power rating.

The system power is often also called net power to distinguish it from the stack power (gross power). The net power, i.e., the system power, is actually utilizable Fahrzeugaufbau for a concrete application. The stack power required for this is approximately 15 % higher because the internal assemblies of the system (particularly the air compressor) must be supplied.

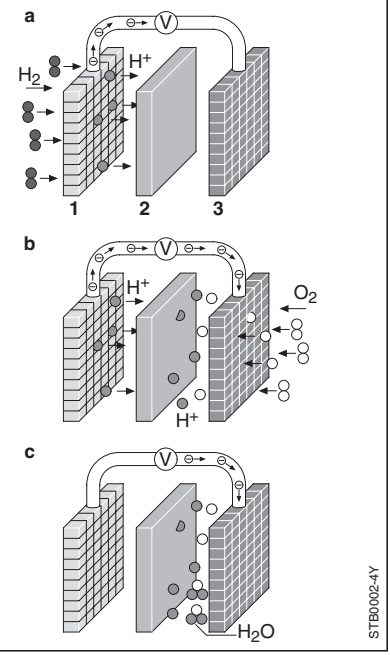
Design and operating principle of the fuel cell

Operating principle

The LT-PEMFC consists of two electrodes (anode and cathode) which are electrically insulated by a polymer-electrolyte membrane (Figure 2). When the anode is supplied with hydrogen (H_2) and the cathode with oxygen (O_2, normally from the air), H_2 and O_2 react with each other even though they are not in direct contact with each other. The reaction is mediated by the membrane and the electric circuit (electron transport, e^-). The following reactions take place:

Anode: $H_2 \rightarrow 2\,H^+ + 2\,e^-$.
Cathode: $O_2 + 4\,H^+ + 4\,e^- \rightarrow 2\,H_2O$.
Overall reaction: $2\,H_2 + O_2 \rightarrow 2\,H_2O$.

Figure 2: Operating principle of the PEM fuel cell
a) Hydrogen oxidation at the anode,
b) Oxygen reduction at the cathode,
c) Water production.
1 Anode, 2 Membrane, 3 Cathode.



In the overall reaction hydrogen reacts with oxygen to produce water, in the course of which electrical energy and heat are generated. The electrical energy can be converted in the circuit to mechanical energy. The waste heat is dissipated via coolant channels.

Figure 3 shows the schematic design of a so-called "cell". Cells have a height of < 1 mm and an area of 200 to 400 cm^2.

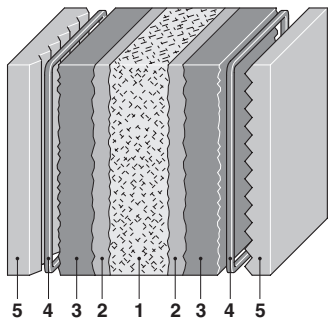
Catalyst layers

The most important components of the fuel cell are the catalyst layers. This is where the chemical reactions which generate the current take place. All the components (including the catalyst layer itself) serve to make available to the catalyst particles within the catalyst layer

– the gaseous reactants (H_2, O_2) in sufficient quantity,
– protons (H^+) from the polymer-electrolyte phase fast enough,
– to offer a conductivity path to electrons,
– to carry off liquid and gaseous product water (H_2O).

Figure 3: Design of the PEM fuel cell

1 Membrane (electrolyte layer),
2 Electrodes (catalyst layer),
3 Gas-diffusion layers,
4 Gaskets,
5 Bipolar plates.



The gas flow begins in the bipolar plates (BPP, metallic or graphitic), which incorporate the channel structures. The gas diffuses via the channels into the gas-diffusion layers (GDL) and from there to the catalyst layer, where it can react (Figure 3).

The electrons flow from the catalyst layer into the gas-diffusion layers, then into the bipolar plates, and from there into the outer circuit (or into the neighboring cell). The protons (H^+) move from the anode through the membrane to the cathode. The product water must be transported from the catalyst layer via the gas-diffusion layers to the channels in the bipolar plate and from there outwards.

Transport mechanisms

Three transport mechanisms occur in the same volume in different phases:

– H^+ through the aqueous component in the polymer-electrolyte membrane,
– gases and H_2O through pores, and
– e^- through a solid-body path.

Making possible all three paths with low resistances places great demands on the material and system design because even the scale at which transport must be guaranteed varies by several orders of magnitude:

– catalyst particles < 15 nm,
– catalyst layer < 30 µm,
– membrane thickness < 30 µm,
– gas-diffusion layer thickness < 300 µm,
– bipolar-plate thickness < 1 mm,
– cell area 200 to 400 cm^2,
– stack height < 1 m.

Electrical properties of the fuel cell

Shape of the polarization curve

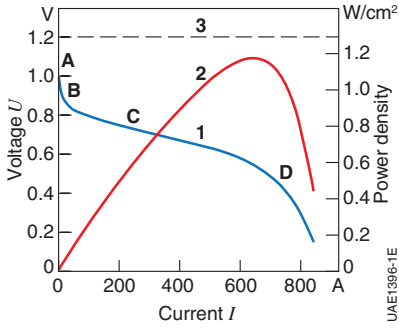
The electrical property of a fuel cell is depicted by the so-called polarization curve (also known as the current-voltage characteristic or U-I characteristic). It generally has the shape depicted in Figure 4.

The voltage of the cell is a direct indicator of its electrical efficiency, where 1.23 V is 100 %. As current and power increase the electrical efficiency of the fuel cell decreases because different losses come into play.

Range A
In the de-energized state (the corresponding voltage is called OCV, open circuit voltage) the theoretical voltage of an H_2-O_2-LT-PEM fuel cell is 1.23 V, which is not achieved however in reality on account of losses (above all H_2 gas-passage losses). These losses cause a low current flow on the inside of the catalyst layer (above all the anode) which cannot be measured towards the outside, which is why the OCV is for the most part in the range of 0.9 to 1 V.

Range B
At low currents the voltage drops quickly first. Here losses which arise due to the chemical reaction of the reactants on the catalyst surface dominate. These losses increase rapidly with small electric currents, but hardly change at high currents (for details on the description of the reaction kinetics as per Butler-Volmer, see [1]). It is to be noted that range A is also actually reduced by this reaction kinetics of the leakage current.

Figure 4: Ranges of the polarization curve (current-voltage characteristic of a cell)
1 Cell voltage,
2 Power density,
3 Theoretical maximum voltage.
A…D Ranges.



Range C
The comparatively linear section is called the ohmic range because the ohmic resistance of the fuel cell is dominant here for the gradient of the curve. The ohmic resistance is made up of the H^+ conductivity of the polymer-electrolyte membrane and the electrical bulk and contact resistances of the fuel-cell layers.

Range D
The inflection of the curve towards 0 V can be called the mass-transport range. The transport of reactants to the catalyst particles is the limiting factor here.

It is to be noted that the losses in the four ranges are always a combination of the losses described above and never is only one loss type (except for OCV) present.

Influences on the polarization curve
The shape of the polarization curve is dependent on the material property of the fuel cell, e.g., on the
– catalyst activity,
– catalyst load,
– membrane thickness,
– ionomer-exchange capacity,
– porosity of the gas-diffusion layers,
– hydrophobicity of the gas-diffusion layers,
– conductivity of the bipolar plates.

Even the operating method has a decisive influence on the shape of the polarization curve, e.g.,
– gas pressures,
– temperature,
– moisture,
– volumetric flow and stoichiometry.

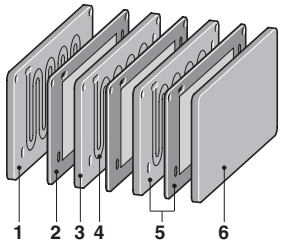
The parameter set under which the polarization curve was plotted is decisive (above all the pressure) and should be observed without fail in a comparison of different fuel-cell types.

Fuel-cell stack
In order to achieve the high voltages required for a technical application, individual cells are electrically connected in series to form a stack (Figure 5). Each cell must be supplied separately with H_2, air, and coolant.

These fuel-cells consist, depending on the requirement, of up to 450 or more cells so that the maximum operating voltages can be up to 450 V.

Typically stacks in the power range up to 120 kW are used for passenger-car drives. High electric currents are achieved either by large membrane surfaces or improved fuel-cell materials (e.g., higher catalyst activities, better mass transport). Currents up to 450 A flow at full load automotive applications.

Figure 5: Design of a fuel-cell stack
H_2 and O_2 are supplied to each cell via gas channels.
1 End plate,
2 Membrane-electrode unit with membrane, catalyst material, and gaskets,
3 Bipolar plate,
4 Gas channels,
5 Repeating units,
6 End plate.



Functioning principle of the fuel-cell system

At the heart of the fuel-cell system is the fuel-cell stack. Subsystems for oxygen and hydrogen supply and for temperature control (Figure 6) are required to operate the stack. There are, in principle, different ways of realizing these subsystems. The one described here shows only one possible variant with an exemplary fuel-cell system power rating of 100 kW net.

Hydrogen supply
The hydrogen is stored in one or more high-pressure tanks at a pressure of around 700 bar. According to the state of the art, type-IV composite tanks with carbon-fiber casing and HDPE lining (high-density polyethylene) are used as an H_2 diffusion barrier. The hydrogen is initially relaxed via a pressure reducer to a mean pressure (approximately 15 bar) and supplied via a hydrogen-metering valve to the anode at low pressure (up to 3 bar).

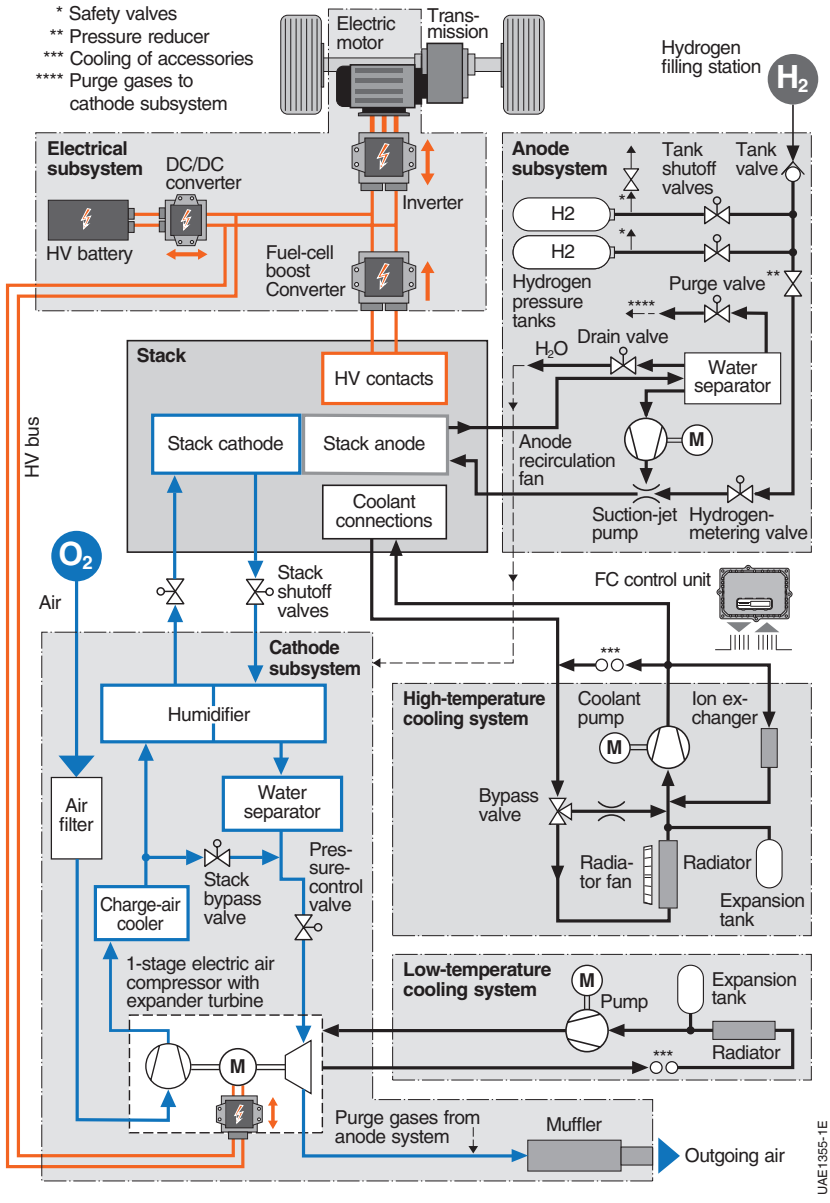
The hydrogen-metering valve is an electrically actuated valve used to set the pressure of the hydrogen on the anode side. In contrast to fuel injectors in internal-combustion engines, the metering valve should set continuous mass flows. The metering valve can also be operated in clocked mode for small pressure and mass-flow pulses, for example when using a suction-jet pump.

A typical throughflow value for 100 kW system power is approximately 1.9 g/s hydrogen.

In order to supply the stack over the entire active surface area with sufficiently high hydrogen concentration, the anode is operated under leaner than stoichiometric conditions (i.e., with H_2 surfeit). To increase system efficiency, the unconsumed hydrogen is returned from the anode output to the input.

Recirculation is effected by a suction-jet pump or a recirculation fan or by a combination of both components. The suction-jet pump, which operates passively, is effective at high power and thus connected with high H_2 mass flow by the hydrogen-metering valve driving the suction-jet pump. The actively actuated recir-

Figure 6: Fuel-cell system with drive



culation fan, however, facilitates recirculation even if no or only minimal power is requested by the system.

Recirculation creates a closed system in which disruptive foreign gases (nitrogen, water vapor) are enriched over time and diffuse from the cathode side to the anode. These foreign gases are regularly discharged via an electrically actuated valve, the purge valve. This valve is mounted on the anode side on the stack outlet.

A water separator permanently removes liquid water from the anode circuit. The water collected in it is regularly drained via a valve. A combination of purge and drain valves in one component is possible.

The hydrogen inevitably dissipated during draining and purging is either strongly diluted with air (when the discharged anode gas is metered into the exhaust pipe) or catalytically converted to water (when metered into the air flow ahead of the cathode). Because this involves only small amounts of H_2 which are not used in the fuel cell to produce electrical energy, the overall efficiency is only slightly reduced.

Oxygen supply

The oxygen required for the electrochemical reaction is taken from the ambient air. The air mass of up to 120 g/s requested for 100 kW system power is inducted by an air compressor, compressed to max. 3 bar, and fed to the cathode side of the fuel cell. The pressure in the fuel cell is adjusted by a dynamic-pressure control valve in the exhaust-gas path downstream of the fuel cell.

To prevent the fuel-cell stack from drying out on the inlet side, the air can be pre-humidified, e.g., via a membrane humidifier, which removes some of the water vapor from the fuel-cell waste air and transfers it to the dry supply air.

In the variant shown the air is diverted via an exhaust-gas turbine seated on the shaft of the electric air compressor. The advantage of this is that using the enthalpy of the waste air relieves the load on the compressor drive motor and increases system efficiency. On the downside, an additional water separator is needed to protect the exhaust-gas turbine against impingement.

Thermal management

Fuel cells have an electrical efficiency in the range of around 50 % (at full load) to 70 % (at part load). This means that a large proportion of heat is released during the conversion of chemical energy to electrical energy which must be dissipated. In normal operation PEM fuel cells typically operate at an operating temperature of 60 °C to approximately 80 °C and thus at a much lower temperature level than internal-combustion engines. The radiator and radiator fan must therefore be designed to be bigger in fuel-cell systems in the vehicle drive in spite of the higher efficiency.

Because the coolant used is in direct contact with the fuel cells, it must not be electrically conducting as otherwise internal short circuit would result. The coolant based on a deionized water-glycol mixture is constantly deonized in the vehicle by an ion exchanger. The setpoint value for coolant conductivity is less than 5 µS/cm.

The electric coolant pump directs a flow of coolant through all the incorporated components. For 100 kW system power the coolant delivery rate can be up to 220 *l*/min. A coolant control valve splits the flow of coolant between the radiator and the radiator bypass.

Sensor systems

Operation of the fuel-cell systems currently requires numerous sensors in the anode path, in the air path, and in the cooling circuit of the system. The pressure sensors familiar from the internal-combustion engine must be adapted to the heightened requirements in the fuel-cell system. This concerns above all robustness against hydrogen and high moisture.

Sensors are furthermore required at numerous points to determine the hydrogen concentration. For fuel-cell systems specifically hydrogen-concentration sensors (see Hydrogen sensors) and coolant-conductivity sensors are required. Furthermore, sensors in the hydrogen and air paths are subject to a heightened requirement with regard to hydrogen resistance and resistance to deionized water.

Safety

H_2-specific dangers

H_2-specific dangers result from the wide combustibility range of hydrogen-air mixtures and the high energy content that is chemically stored in hydrogen. Depending on how the hydrogen is stored, even the high pressure or the low temperature can constitute a danger. The 700 bar compressed-gas tank has established itself as the standard for “on-road fuel tanks” and the associated infrastructure.

The handling of hydrogen in motor vehicles is governed by statutory provisions such as for example UNECE R134 [2], EU406/2010 [3], UNECE GTR13 [4], and applicable standards such as ISO 12619 [5] or ANSI HGV 3.1 [6]. Here the primary focus of the statutory provisions is on sealing the hydrogen-carrying parts, particularly the tank.

With regard to use in vehicles, it is essential to prevent the uncontrolled release and accumulation of hydrogen in closed or partly closed rooms. In addition, a design goal is in place to prevent ignition sources in the immediate vicinity of hydrogen-carrying parts. Measures for detecting unwanted leakages are either system-internal diagnostics or gas sensors installed in suitable positions.

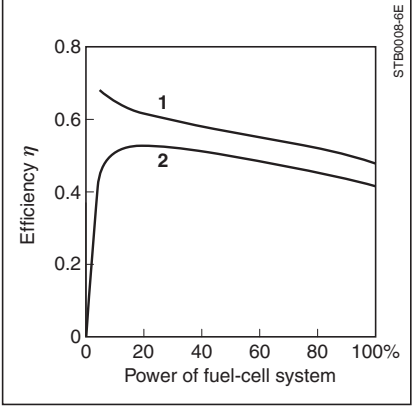
Efficiency of the fuel-cell system

In addition to quickly providing the requested power to the stack under the best possible operating conditions, it is important to design the system for primarily operating at high efficiency.

In Figure 7, the efficiency of a fuel-cell stack is compared by way of example with that of a fuel-cell system. The secondary loads/consumers (e.g., air compressor, coolant pump) take up some of the electric power and thereby reduce the overall efficiency of the system. Fuel-cell systems demonstrate substantially higher efficiency than an internal-combustion engine, particularly in the frequently used part-load range.

Figure 7: Efficiency of fuel-cell stack and fuel-cell system

1 Efficiency of fuel-cell stack,
2 Efficiency of fuel-cell system.



Fuel cell in the drivetrain

Figure 8 depicts the drivetrain of a truck with a fuel cell.

Electrical-system typologies
One or more DC/DC converters (at least one DC/DC converter is required) spread the power between the fuel-cell system, the traction battery, and the electric drive. In the case of a topology with two DC/DC converters – one for the fuel-cell stack and one for the traction battery – an optimized voltage can be set independently of both. The latter offers advantages for the design of the HV system components (high voltage) such as the compressor.

For safety reasons, the traction electrical system is isolated from the vehicle ground.

Components of the drivetrain
Electric drive
The electric drive consists of power electronics (an inverter) and a synchronous or asynchronous machine that is energized by the inverter to generate the required engine torque (see Electrical machines for motor-vehicle propulsion). Because the electric drive has a high power rating (approximately 150 kW), it is operated at voltages of up to 450 V. A voltage of up to 800 V can be selected for powerful vehicle classes and trucks.

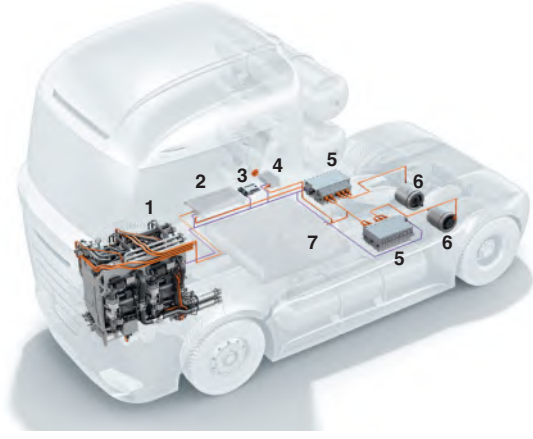
The direct voltage of the traction electrical system is converted in the inverter into a multi-phase alternating current (see Inverter) where the amplitude is regulated as a function of the desired drive torque.

During braking the electric machine is switched to generator mode. In this process the vehicle's kinetic energy is converted into electrical energy and fed back to the traction battery.

Traction battery
Depending on the application, high-capacity or high-energy batteries are used with voltages of between 150 V and 400 V (see Batteries for electric and hybrid drives). Nickel-metal-hydride or lithium-ion batteries are used in high-capacity applications, while only lithium-ion batteries are used in high-energy applications. A battery-management system monitors the state of charge and the capacity of the battery.

DC/DC converter for the traction battery
A DC/DC converter regulates the traction battery's charge and discharge currents, where maximum currents of up to 300 A flow through the battery. This converter may be omitted in certain system configurations.

Figure 8: Electrically powered truck with fuel cell
1 Fuel-cell system,
2 Fuel-cell DC/DC converter,
3 Vehicle control unit (VCU),
4 Battery charging socket,
5 Inverter,
6 e-machines,
7 HV battery.



DC/DC converter for the fuel cell
Another DC/DC converter regulates the current from the fuel-cell stack, where maximum currents of up to 500 A flow through the stack. This converter too may be omitted in certain system configurations.

DC/DC converter for 12V vehicle electrical system
There is also – as in conventional vehicles – a 12V electrical system for the low-power electric loads/consumers. The 12V vehicle electrical system is supplied from the traction electrical system. A DC/DC converter is used between the two networks for this purpose. This converter is electrically isolated for safety reasons. It operates unidirectionally or bidirectionally with a power output of up to 3 kW to keep the currents in the 12V electrical system within the limits.

Outlook
Drives with fuel cells have already proven that they are suitable for everyday use. Estimates from 2015 by the Department of Energy ([7], [8]) start out from fuel-cell-system costs of 54 US$ per kW at 80 kW and 500,000 vehicles per year and additionally 488 US$ per kilogram of hydrogen for the tank.

System simplifications deliver both cost and reliability advantages. One approach is to develop new polymer membranes for fuel cells where it is not necessary to humidify the reaction gases and which at the same time allows the operating temperature to be increased.

Passenger-car models available on the market are the Hyundai Nexo, the Toyota Mirai, the Honda Clarity Fuel Cell, and the Mercedes-Benz GLC F-Cell (as at 2020). German automobile manufacturers are planning market launches in the coming years.

References
[1] F. Barbir: PEM Fuel Cells Theory and Practice. Elsevier Inc., 2013.
[2] UNECE R134: Regulation No 134 of the Economic Commission for Europe of the United Nations (UN/ECE) – Uniform provisions concerning the approval of motor vehicles and their components with regard to the safety-related performance of hydrogen-fuelled vehicles (HFCV) [2019/795].
[3], EU 406/2010: Commission Regulation (EU) No 406/2010 of 26 April 2010 implementing Regulation (EC) No 79/2009 of the European Parliament and of the Council on type-approval of hydrogen-powered motor vehicles.
[4], UNECE GTR13: Global Technical Regulation concerning the hydrogen and fuel cell vehicles.
[5] ISO 12619: Road vehicles – Compressed gaseous hydrogen (CGH2) and hydrogen/natural gas blend fuel system components.
[6] ANSI HGV 3.1: Fuel system components for compressed hydrogen gas powered vehicles.
[7] J. Marcinkoski, J. Spendelow, A. Wilson, D. Papageorgopoulos: DOE Hydrogen and Fuel Cells Program Record #15015 Fuel Cell System Cost.
https://www.hydrogen.energy.gov/pdfs/15015_fuel_cell_system_cost_2015.pdf
[8] B. James: Onboard Type IV Compressed Hydrogen Storage System Cost Analysis.
https://energy.gov/sites/prod/files/2016/03/f30/fcto_webinarslides_compressed_h2_storage_system_cost_022516.pdf.

Electromobility for two-wheelers

Features

Definition and delimitation

Like electromobility in the four-wheel sector, mobility in the two-wheel sector is also following the trend of electrification. With the focus on purely electric vehicles, vehicles in this sector can be a PHEV (Plug-in Hybrid Electric Vehicle), an HEV (Hybrid Electric Vehicle) or a BEV (Battery Electric Vehicle). Only vehicles with a battery-electric drivetrain will be considered in more detail in the following.

EC168/2013

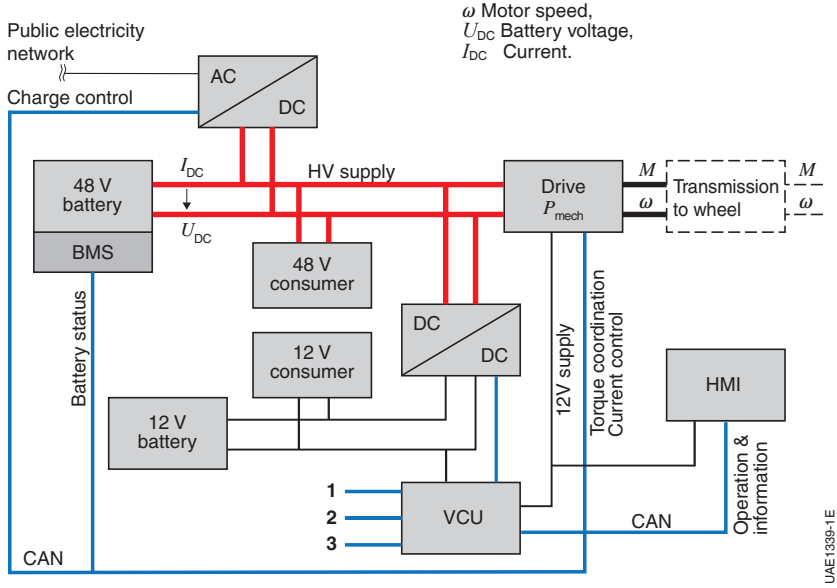
It is possible with reference to the classification of light vehicles in accordance with EC168/2013 [1] to categorize in particular two- and three-wheel vehicles into vehicle segments. The vehicle classes for four-wheel light vehicles will not be considered here. The segmentation applies to vehicles with internal-combustion engines and electrically powered vehicles.

Requirements

Two-wheel electric vehicles are characterized by the fact that they have, compared with passenger cars, energy accumulators with relatively low energy content (in kWh). There is here typically a P/E ratio (power to energy, see Batteries for electric and hybrid drives) of 0.5 to 5.0 kW/kWh for the segment of electrically powered two-wheel vehicles. Moreover, passive cooling methods (primarily air cooling) are used in such vehicles in the lower power range below 11 kW. Two-wheel vehicles are sometimes subject to statutory provisions for antilock braking systems (e.g., in the EU from 11 kW).

Figure 1: System architecture

1 Twist grip, 2 Brake lever, 3 Riding switch.
BMS Battery Management System, VCU Vehicle Control Unit,
AC Alternating Current, DC Direct Current,
HMI Human Machine Interface (operation and display), CAN Controller Area Network,
M Drive torque,
ω Motor speed,
U_{DC} Battery voltage,
I_{DC} Current.



System architecture

Aside from chassis, steering, suspension, and damping, the following main components are used in an electric two-wheeler (Figure 1): driving motor with power electronics, DC/DC converter, charger, battery, vehicle control unit (VCU), and display instrument. Figure 1 also shows the networking of the individual main components of the electric drivetrain (traction energy via 48 V electrical system, communication paths via CAN bus).

The transmission of electrical-mechanical energy from the electric machine to the wheel can be achieved by means of different technical solutions. The connection to the driving wheel can be established via a transmission, a toothed belt or a chain.

Figure 1 is also valid for operating voltages greater than 60 V, but some benefits of the 48 V technology are lost. These are on the one hand the omission of technical protection measures (shock protection); on the other hand, untrained employees (non-HV-technical specialists) are permitted to work on the vehicle. Significant cost savings are thus possible.

Motor arrangement
Two essentially different motor arrangements may be used within a system architecture. The central motor (Figure 2a) is characterized in conjunction with its power transmission by higher torques. The wheel-hub motor has the advantage of lower noise emissions (e.g., due to the absence of a chain).

Battery arrangement
Electric two-wheels can furthermore be distinguished on the basis of the installation position of the battery. The battery can be installed close to the steering column at the front of the vehicle, in the running-board area (Figure 2c) or in the former helmet compartment under the seat in the vehicle (Figure 2b).

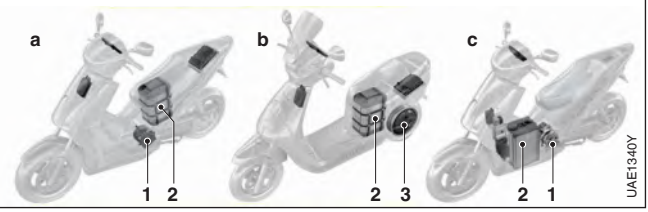
In order to establish the installation position of the battery, it is necessary to define beforehand the use to which the vehicle will be put. The type of battery used depends on this intended use. This constitutes an important extra dimension in the definition of electric two-wheelers. A distinction is made here between permanently installed batteries, portable batteries with an additional plug-in process, and portable batteries without an additional plug-in process.

Charging methods
Likewise, different charging methods may be used for an electrified two-wheeler. Generally, the following different methods may be used:
– charging with a permanently installed charger,
– charging with a portable (mobile) charger or
– charging via swapping systems.

Here too the intended use of the vehicle is crucial in determining which charging method is the most expedient.

References
[1] EC168/2013: Regulation (EU) No. 168/2013 of the European Parliament and of the Council of 15 January 2013 on the approval and market surveillance of two- or three-wheel vehicles and quadricycles.

Figure 2: Arrangement of the electric drive system
a) Central motor,
b) Battery in helmet compartment,
c) Battery in running-board area.
1 Central drive,
2 Battery,
3 Wheel-hub drive.



Emission-control legislation

Provided for information purposes only, no liability assumed for completeness!

Overview

California introduced the first emission-control legislation for gasoline engines in the mid-1960s. These regulations became progressively more stringent in the ensuing years. In the meantime, all industrialized countries have introduced emission-control laws which define limits for gasoline and diesel engines, as well as the test procedures employed to confirm compliance.

In many countries, regulations governing exhaust-gas emissions are supplemented by limits on evaporative losses from the fuel systems of vehicles with gasoline engines. Statutory OBD is addressed in the chapter "Diagnostics".

The most important legal restrictions on exhaust-gas emissions are listed below:
- CARB legislation (California Air Resources Board)
- EPA legislation (Environmental Protection Agency), USA
- EU legislation (European Union) and the corresponding UN/ECE regulations (United Nations/Economic Commission for Europe)
- Japanese legislation
- Chinese legislation

Classification
Countries with legal limits on motor-vehicle emissions divide vehicles into various classes:
- Passenger cars: Emission testing is conducted on a chassis dynamometer.
- Light commercial vehicles: Depending on national legislation, the top limit for gross vehicle weight (GVW) is 3.5 to 6.35 t. Emission testing is performed on a vehicle chassis dynamometer (as for passenger cars)
- Heavy commercial vehicles: Gross vehicle weight over 3.5 to 6.35 t (depending on national legislation). Emission testing is performed on an engine test bench.
- Non-road (e.g., construction, agricultural, and forestry vehicles): Emission testing is performed on an engine test bench, as for heavy commercial vehicles.

Furthermore, there are emission regulations for other two-, three-, and four-wheel vehicles (see section Emission legislation for motorcycles), for locomotives, for boats and ships, and for non-road mobile machinery.

Test procedures
Japan and the European Union have followed the lead of the United States by defining test procedures for certifying compliance with emission limits. These procedures have been adopted in modified or unrevised form by other countries.

Legal requirements prescribe any of three different tests, depending on vehicle class and test objective:
- Type approval (TA) to obtain General Certification,
- Random testing of vehicles from serial production conducted by the approval authorities (COP, Conformity of Production),
- In-field monitoring for checking the emission-reduction systems of privately owned production vehicles under real driving conditions.

Type approval
Type approvals are a precondition for granting General Certification for an engine or vehicle type. For this purpose, test cycles must be driven under specific operating conditions and emission limits must be complied with. The test cycles and emission limits are specified individually by each nation.

Dynamic test cycles are specified for passenger cars and light commercial vehicles. The country-specific differences between the two procedures are rooted in their respective origins (see Test cycles for passenger cars and light-duty trucks):
- Test cycles designed to mirror conditions recorded in actual highway operation, e.g., the FTP test cycle (Federal Test Procedure) in the USA or the UN ECE WLTC (World Light-duty Test Cycles),

– Synthetically generated test cycles consisting of phases at constant cruising speed and acceleration rates, e.g., the MNEDC (Modified New European Driving Cycle) of the European Union.

The mass of pollutant emissions from each vehicle is determined by operating it in conformity with speed cycles precisely defined for the test cycle. During the test cycle, the exhaust gases are collected for subsequent analysis to determine the pollutant mass emitted during the driving cycle (see Exhaust-gas measuring techniques).

For heavy-duty trucks and non-road applications, steady-state (e.g., 13-stage test) and dynamic test cycles (e.g., US HDDTC or ETC) are carried out on the engine test bench (see Test cycles for heavy commercial vehicles).

Serial testing (Conformity of Production)
Vehicle manufacturers normally perform serial testing themselves during production as part of quality control. The same test procedures and the same limits are generally applied as for type approval. The approval agency audits the serial testing and can request retests. The strictest requirements are applied in the USA where very nearly 100 % quality monitoring is required (EPA: "Compliance Assurance Program" CAP; CARB: "Production Vehicle Evaluation" PVE).

In-field monitoring
In-field monitoring involves the detection of type-specific defects (e.g., design or manufacturing defects, defective maintenance instructions) which result in dramatically increased pollutant emissions during the operation of vehicles under normal conditions of use. To this end, a check is carried out in the field on production vehicles to ascertain that they comply 100 % with the emission regulations. Privately owned production vehicles are randomly chosen for testing. The mileage and age of the vehicle must be within defined limits.

Emission-control legislation for passenger cars and light commercial vehicles

USA CARB legislation
The emission limits of CARB (California Air Resources Board) for passenger cars (PC) and light commercial vehicles (light-duty trucks; LDT) are specified in the emission-control standards LEV I, LEV II, and LEV III (LEV, Low Emission Vehicle, i.e., vehicles with low exhaust and evaporative emissions). Since model year 2004 the LEV II standard has applied to all new vehicles up to a gross vehicle weight of 14,000 lbs (lb: pound; 1 lb = 0.454 kg, 14,000 lb = 6.35 t). Model year 2015 marked the beginning of the incremental introduction of the new emission level LEV III to be completed by 2025 (or 2022 for evaporative emissions).

Originally the CARB legislation only applied in the US state of California, but has since been adopted in other states.

Vehicle classes
Figure 1 shows an overview of the categorization of vehicles into vehicle classes.

Emission limits
CARB legislation specifies limits for the carbon monoxide (CO), nitrogen oxides (NO_x), non-methane organic gases (NMOG), formaldehyde (HCHO), and particulate matter (diesel: LEV I to LEV III; gasoline: LEV III only) (Figure 2). On account of the very low emission limits, the LEV III legislation provides for an aggregate value of NMOG and NO_x.

Pollutant emissions are measured in the FTP 75 driving schedule (Federal Test Procedure). The emission limits are correlated with the route driven during the test and are expressed in grams per mile.

The SFTP (Supplemental Federal Test Procedure) standard, comprising two further test cycles (SC03 and US06 cycles) was introduced in 2001. There are also further limits that require compliance in addition to FTP emission limits.

Exhaust-gas categories
Automotive manufacturers are at liberty to deploy a variety of vehicle concepts within the permitted limits, providing they maintain a fleet average (see the section entitled “Fleet average”). The concepts are allocated to the following exhaust-gas categories, depending on their emission values for NMOG, CO, NO_x, and particulate emissions:
– LEV (Low-Emission Vehicle),
– ULEV (Ultra-Low-Emission Vehicle),
– SULEV (Super Ultra-Low-Emission Vehicle).

Since 2004 new vehicle registrations have been governed by the LEV II exhaust-emission standard. The CO and NMOG limits from LEV I remain unchanged, but the NO_x limit is substantially lower for LEV II.

LEV III provides for a total of six vehicle categories (Figure 2), one of which is below SULEV. In addition to the categories of the LEV emission-control standards, three categories of zero-emission or partial zero-emission vehicles are defined (see ZEV program).

Figure 1: Vehicle classes, CARB legislation
LDT Light Duty Truck, MDV Medium Duty Vehicle, HDV Heavy Duty Vehicle, PC Passenger Car, LDV Light Duty Vehicle, LVW Loaded Vehicle Weight (vehicle tare weight plus 300 lbs), GVW Gross Vehicle Weight.

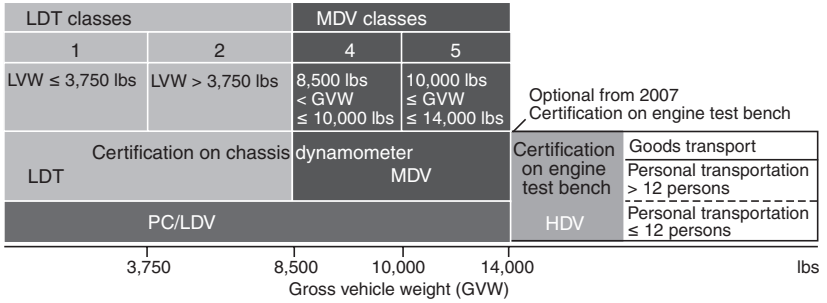
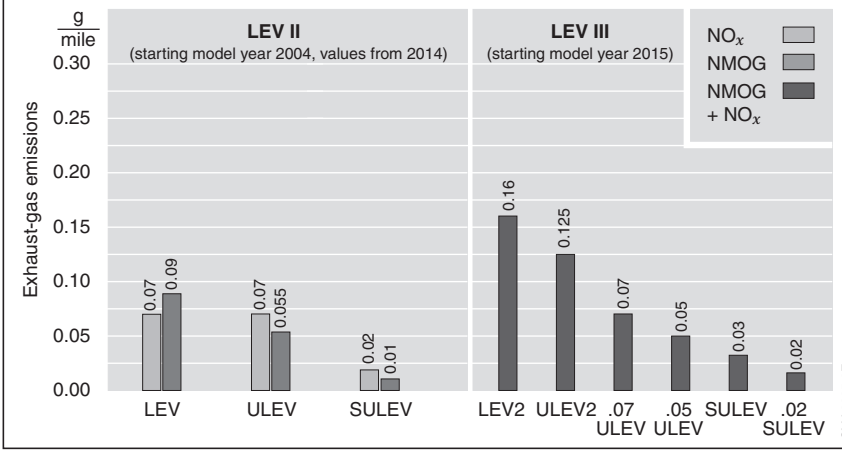


Figure 2: Exhaust-gas categories and emission limits for NO_x and NMOG of CARB legislation for passenger cars and light commercial vehicles



Phase-in
The LEV emission-control standards will not be introduced from one year to the next, but instead by means of a "phase-in", i.e., the incremental introduction of the requirements over a number of years for ever larger proportions of the new-car fleet, e.g. for LEV II 25%/50%/75%/100% of the newly registered vehicles in the model years 2004/5/6/7. This is accompanied by a "phase-out" of the previous regulations.

The LEV III standard will be introduced between 2015 and 2025, phase-out of LEV II by 2019. Particulate emissions are covered by separate phase-in and phase-out from 2017 onwards.

Evaporative emissions
The subject of evaporative emissions is addressed in the chapter "Exhaust-gas measuring techniques".

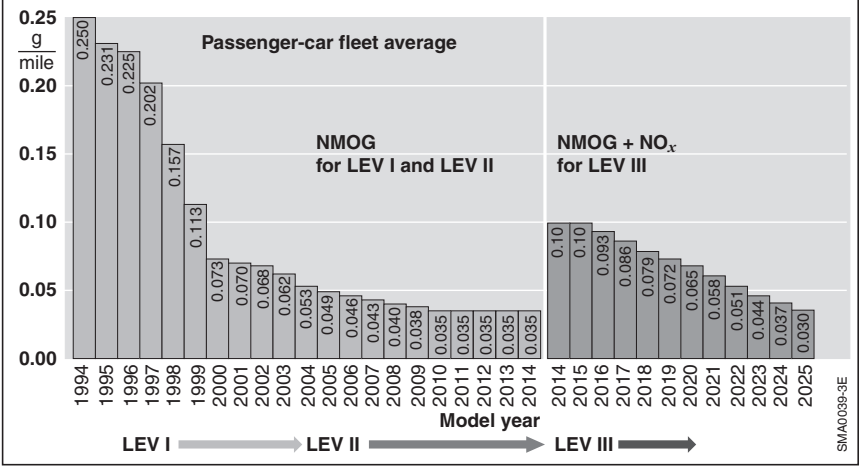
Durability
In order to gain approval for vehicle types, vehicle manufacturers must certify that the emissions of limited pollutants do not exceed the respective limits over 50,000 miles or 5 years ("half useful life") and 100,000 miles (LEV I) or 120,000 miles (LEV II) or 10 years ("full useful life"). LEV III has extended durability to 150,000 miles.

Durability test
The vehicle manufacturer must supply two vehicle fleets from its production line for the durability test. One fleet in which each vehicle has driven 4,000 miles before the test, and one fleet for the endurance test with which the deterioration factors of individual pollutant components are measured.

Endurance testing entails subjecting the vehicles to specific driving cycles over distances of 100,000 and 120,000 or 150,000 miles. Exhaust-gas emissions are tested at intervals of 5,000 miles and may not exceed the limits. Servicing and maintenance tasks may only be performed at the specified intervals.

Fleet average
Each vehicle manufacturer must ensure that its vehicles do not exceed an average specific emission limit for exhaust-gas emissions (Figure 3). The NMOG and NO_x emissions are used as the criteria in this regard with LEV III (for LEV I/II only NMOG). The fleet average is calculated for LEV I and LEV II from the average of the NMOG limit for "half useful life" produced by all of the manufacturer's vehi-

Figure 3: Passenger-car fleet averages in comparison with NMOG standard



cles are sold within one year. The limits for "full useful life" apply to LEV III. The emission limits for the fleet average are different for passenger cars and light commercial vehicles.

The emission limit for the NMOG fleet average is reduced every year. To meet the lower fleet limit, manufacturers must produce progressively cleaner vehicles in the more stringent emissions categories in each consecutive year.

Fleet consumption (fuel consumption)
US federal lawmakers specify mandatory requirements on vehicle manufacturers with regard to the mean fuel consumption of their vehicle fleets, or the number of miles driven per gallon (federal law, applicable in all states, the relevant authority is the "National Highway Traffic Safety Administration", NHTSA). New vehicles registered in the area of responsibility of CARB are governed by the same CAFE regulations for determining fleet fuel consumption as for EPA (see EPA legislation, section "Fleet consumption").

Emission-free vehicles
The ZEV program ("Zero-Emission Vehicle") is to be used to force the development and market introduction of "zero-emission vehicles" ("technology forcing"), which are also intended to help reduce greenhouse emissions.

The ZEV program defines three categories of zero-emission and partial zero-emission vehicles. Genuine ZEV vehicles may not release any emissions when they are in operation. These vehicles are electric cars which are operated by a battery or a fuel cell.

PZEVs (Partial Zero-Emission Vehicles) are not emission-free, but they do emit particularly low levels of pollutants. They are weighted using a factor of more than 0.2, depending on the emission-limit standard. The following requirements must be satisfied for the minimum factor of 0.2:
– SULEV certification for a durability of 150,000 miles or 15 years,
– Warranty coverage extending over 150,000 miles or 15 years on all emission-related components,
– No evaporative emissions from the fuel system (0 EVAP, zero evaporation). This is achieved by extensive encapsulation of the tank and fuel system. This results in greatly reduced evaporative emissions by the overall vehicle.

AT PZEVs (Advanced Technology PZEVs) are hybrid vehicles with a gasoline or diesel engine and electric motor and gas-powered vehicles (operation with compressed natural gas, with hydrogen).

The ZEV program envisages for the major automobile manufacturers minimum quantities for ZEV, AT PZEV and PZEV vehicles which will increase gradually between 2005 and 2017 and provide for ever greater proportions of AT PZEVs and ZEVs. The quantities are not calculated directly, but instead by means of so-called "ZEV credits", which depend among other things on the technology used, their performance, and the model year. From 2018 the PZEV category will be dropped and the necessary quantities of AT PZEVs (now transitionally TZEVs) and ZEVs will rise dramatically by 2025.

The relevant figures for the ZEV, AT-PZEV and PZEV emission-limit categories are 150,000 miles or 15 years ("full useful life").

In-field monitoring
Non-routine inspection
Random emission testing is conducted on in-use vehicles using the FTP 75 test cycle and – for vehicles with gasoline engines – an evaporative-emission test. Depending on the relevant exhaust-gas category, vehicles with mileage readings below 90,000 or 112,500 miles are tested.

Vehicle monitoring by the manufacturer
Official reporting of claims or damage to specific emissions-related components and systems has been mandatory for vehicle manufacturers since model year 1990. The reporting obligation remains in force for a maximum period of 15 years, or 150,000 miles, depending on the warranty period applying to the component or assembly.

The reporting method is divided into three stages with an incremental amount of detail: Emissions Warranty Information Report (EWIR), Field Information Report (FIR), and Emission Information Report (EIR). The California Air Resources Board is notified of information regarding complaints, fault quotas, fault analysis, and impacts on emissions. The agency uses the Field Information Report as the basis to decide whether to enforce a recall action on the vehicle manufacturer.

USA EPA legislation

EPA (Environmental Protection Agency) legislation applies to all of the states where the more stringent CARB stipulations from California are not in force.

The EPA legislation has been covered by the Tier 2 emission-control standard since 2004. The new Tier 3 emission-control standard will be introduced incrementally between 2017 and 2025. Tier 3 achieves an extensive harmonization of the requirements with the Californian LEV III program, e.g., the same certification categories, albeit with its own designation ("bin").

Vehicle classes

With the transition to Tier 2, an additional vehicle class has been introduced in the form of the MDPV (Medium-Duty Passenger Vehicle) (Figure 4). Thus, all vehicles up to a permissible total weight of 10,000 lbs (4.54 t) which are designed to carry up to 12 persons, are certified on a chassis dynamometer.

Light commercial vehicles are divided into two groups: LLDTs (Light Light-Duty Trucks) with a gross vehicle weight of up to 6,000 lbs (2.72 t) and heavier HLDTs (Heavy Light-Duty Trucks) with a gross vehicle weight of up to 8,500 lbs (3.86 t).

Since 2007 certification on a chassis dynamometer has also been optionally possible for vehicles weighing up to 14,000 lbs (6.35 t).

The classification of motor vehicles into the vehicle classes defined in Tier 2 has been retained for Tier 3.

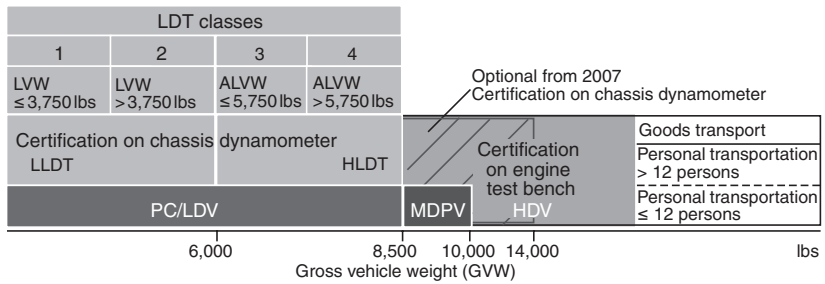
Emission limits

EPA legislation specifies limits for the pollutants carbon monoxide (CO), nitrogen oxides (NO_x), non-methane organic gases (NMOG), formaldehyde (HCHO), and particulate matter (PM). Pollutant emissions are measured in the FTP 75 driving schedule. The emission limits are correlated with the route driven during the test and are expressed in grams per mile.

The SFTP (Supplemental Federal Test Procedure) standard, comprising two further test cycles (SC03 and US06 cycles) has been in force since 2000. Prevailing emission limits must be met in addition to FTP emission limits.

Figure 4: Vehicle classes, EPA legislation

LDT Light Duty Truck, MDV Medium Duty Vehicle, HDV Heavy Duty Vehicle, PC Passenger Car, LLDT Light Light-Duty Truck, HLDT Heavy Light-Duty Truck, MDPV Medium-Duty Passenger Vehicle, LDV Light Duty Vehicle, LVW Loaded Vehicle Weight, GVW Gross Vehicle Weight, ALV Adjusted Loaded Vehicle Weight (0.5 × tare weight + 0.5 × gross weight).



Since the introduction of Tier 2 standards in 2004, vehicles with diesel and gasoline engines have been subject to identical exhaust-emission limits.

Exhaust-gas categories

Tier 2

For Tier 2, the limits were divided into 10 (for LDV and LLDT) and into 11 (for HLDT and MDPV) emission standards (bins) (Figure 5). For LDV and LLDT, Bin 9 and Bin 10 ceased to apply in 2007, and, for HLDT and MDPV, Bin 9 through Bin 11 ceased to apply in 2009.

The transition to Tier 2 has produced the following changes:

– Introduction of fleet averages for NO_x,
– Formaldehyde (HCHO) is subject to a separate pollutant category,
– LDV and LLDT are treated identically to the greatest possible extent with regard to FTP limits,
– "Full useful life" is increased, depending on the emission standard (Bin), to 120,000 or 150,000 miles.

Tier 3

Seven selectable certification bins will furthermore be made available for Tier 3. However, the standards are now formulated both for the fleet values and for the bins as an aggregate value of NMOG and NO_x.

The Tier 3 legislation closely follows the Californian LEV III legislation in order to simplify the job of certification in the EPA and CARB areas for the vehicle manufacturers.

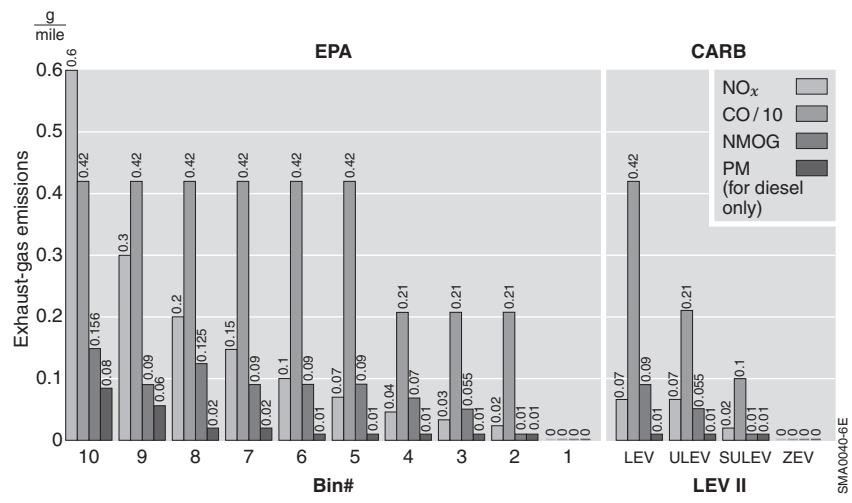
Durability has been extended for Tier 3 to 150,000 miles (optional, alternatively 120,000 miles with stricter limits values) or 15 years compared with 100,000 miles or 10 years for Tier 2.

Tier 3 also covers heavy pick-ups and vans with a gross vehicle weight in excess of 6,500 lbs, which are now driven on a chassis dynamometer including the demanding SFTP test cycles. Within the Tier 2 legislation these vehicles were excluded from the SFTP tests.

Particulate-emission limits are prescribed with the Tier 3 legislation for each vehicle. A limit for the particulate number (PN) is currently not provided for.

The Tier 3 legislation is governed by the same fleet values for NMOG + NO_x as for LEV III.

Figure 5: Tier 2 emission limits for EPA legislation compared with CARB emission limits for LEV II



Phase-in
At least 25 % of all new passenger-car and LLDT registrations were required to conform to Tier 2 standards which took effect in 2004. The phase-in rule stipulated that an additional 25 % of vehicles would be required to conform to the Tier 2 standards in each consecutive year. All vehicles have been required to conform to Tier 2 standards since 2007. The phase-in period for HLDT and MDPV terminated in 2009.

Tier 3 will be introduced between 2017 and 2025, phase-out of Tier 2 by 2019. Particulate emissions are covered by separate phase-in and phase-out from 2017 onwards.

Evaporative emissions
The subject of evaporative emissions is addressed in the chapter "Exhaust-gas measuring techniques".

Durability
The same criteria as for CARB apply to durability.

Fleet average
NO_x emissions are used to determine fleet averages for individual manufacturers under the EPA Tier 2 legislation. Up to 2008 the value was 0.2 g/mile, since 2008 it has been 0.07 g/mile. With Tier 3 the NMOG+NO_x fleet average will be introduced as for CARB.

Fleet consumption (fuel consumption)
US federal lawmakers specify mandatory requirements on vehicle manufacturers with regard to the mean fuel consumption of their vehicle fleets, or the number of miles driven per gallon (federal law, applicable in all states, the relevant authority is the "National Highway Traffic Safety Administration", NHTSA). Here it is the metric "miles per gallon of fuel (mpg)" and not as otherwise usual in Europe for example the "amount of fuel used per unit distance" that is used. This representation of "fuel economy" corresponds to the reciprocal value of consumption per unit of distance.

The prescribed CAFE value (Corporate Average Fuel Economy) was until 2010 27.5 mpg for passenger cars. This equates to a fuel consumption of 8.55 *l* of gasoline per 100 km. Up to 2004 the value for light commercial vehicles was 20.7 mpg or 11.36 *l* of gasoline per 100 km and was increased by 2010 to 23.5 mpg.

For model year 2011 the CAFE system was restructured for passenger cars and light-duty trucks (among others the definitions of passenger cars and light trucks) and ambitious target values were set: 33.3 mpg and 22.8 mpg for 2011, 40.1 mpg and 25.4 mpg for 2014, 43.4 mpg and 26.8 mpg for 2016, 46.8 mpg and 33.3 mpg for 2021, and 56.2 mpg and 40.3 mpg for 2025.

For the stages 2012 through 2016 and 2017 through 2025 an EPA regulation harmonized with the CAFE regulation for limiting greenhouse emissions applies in parallel to the CAFE legislation. The target value for a vehicle is dependent on the tire contact area between the wheels ("footprint"). Target values are 34.1 mpg (for CAFE) and 250 g CO_2 equivalent per mile (for EPA) for 2016. In the period from 2017 through 2025 the target values will be further increased in stages, for EPA up to 54.5 mpg (163 g CO_2 equivalent per mile) and for CAFE up to 49.7 mpg (179 g CO_2 equivalent per mile). The values for 2021 through 2025 were suspended in 2018 by the Trump administration. This led to complaints by CARB and many other US states. A reassessment by the Biden administration is expected.

At the end of each year the average "fuel economy" for each vehicle manufacturer is calculated based on the numbers of vehicles sold. The manufacturer must remit a penalty fee of $5.50 per vehicle for each 0.1 mpg its fleet falls short of the target. There is the possibility by way of a "credit/debit system" of balancing overfulfillment and underfulfillment in a year with "credits" and "debits" from other years.

Buyers will also have to pay a gas-guzzler tax on vehicles with especially high fuel consumption. Here, the limit is 22.5 mpg (10.45 *l* per 100 km). These penalties are intended to promote the sale of vehicles offering greater fuel economy.

The FTP 75 test cycle and the highway cycle are applied to measure CAFE fuel economy (see the section entitled "US test cycles").

A Fuel Economy Label provides vehicle buyers with information on fuel consumption. Model year 2008 sees the introduction of "5 cycle fuel economy" (also called "5 cycle method"), which is intended to better reflect real driving conditions. This takes into account measurements in the SFTP schedules and in the FTP at –7 °C, which include among others aggressive acceleration, high final speed, and also operation with air conditioning.

In-field monitoring

Non-routine inspection

EPA legislation, just like CARB legislation, provides for an exhaust-gas emission test in accordance with the FTP 75 test method to be carried out on in-use vehicles on a random-test basis. Testing is conducted on low-mileage vehicles (10,000 miles, roughly one year old) and higher-mileage vehicles (50,000 miles, however at least one vehicle per test group with 90,000 or 105,000 miles, depending on the emission standard; roughly four years old). The number of vehicles is dependent on the number sold. For vehicles with gasoline engines, at least one vehicle per test group is also tested for evaporative emissions.

Vehicle monitoring by the manufacturer

For vehicles after model year 1972, the manufacturer is obliged to make an official report concerning damage to specific emission-related components or systems if at least 25 identical emission-related parts in a model year are defective. The reporting obligation ends five years after the end of the model year. The report comprises a description of damage to the defective component, presentation of the impacts on exhaust-gas emissions, and suitable corrective action by the manufacturer. The environmental authorities use this information as the basis for determining whether to issue recalls to the manufacturer.

EU legislation

The Directives of European emission-control legislation are proposed by the EU Commission and ratified by the Council of Environment Ministers and the EU Parliament. The basis of emission-control legislation for passenger cars and light commercial vehicles is Directive 70/220/EEC [1] from 1970. For the first time it defined exhaust-emission limits, and the provisions have been updated ever since.

The emission limits for passenger cars and light commercial vehicles LCV (light-duty trucks LDT) are contained in the emission-control standards Euro 1 (from 07/1992), Euro 2 (01/1996), Euro 3 (01/2000), Euro 4 (01/2005), Euro 5 (09/2009), and Euro 6 (09/2014).

Instead of a "phase-in" over several years as in the USA, a new emission-control standard is introduced in two stages. In the first stage newly certified vehicle types must adhere to the newly defined emission limits. In the second stage – generally one year later – every newly registered vehicle (i.e., all types) must adhere to the new limits. Legislators can inspect production vehicles for compliance with emission limits (COP, Conformity of Production, and ISC In-service Conformity Check).

EU directives allow tax incentives for vehicles that comply with upcoming exhaust-gas emission standards before they actually become law. Depending on a vehicle's emission standard, there are also a number of different motor-vehicle tax rates in Germany.

Vehicle classes

Until the Euro 4 legislation expired, vehicles with a permissible total weight below 3.5 t were certified on chassis dynamometers; in this respect, a distinction was made between passenger cars (carrying up to nine persons) and light commercial vehicles (LDT) for goods transportation. There are three classes for LDT (Figure 6), depending on the vehicle reference weight (tare weight + 100 kg). Engine certifications are performed for buses (carrying more than nine persons) and for vehicles with a permissible total weight in excess of 3.5 t. LDT engines can

optionally also be certified on an engine test bench.

When the Euro 5 and Euro 6 legislation comes into force, the vehicle reference weight (tare weight + 100 kg) will be the distinguishing criterion with regard to the certification procedure. Vehicles with a reference weight of up to 2.61 t will be certified on a chassis dynamometer. Vehicles with a reference weight in excess of 2.61 t will be certified on an engine test bench. Flexible variations will be possible.

Emission limits

The EU standards specify limits for carbon monoxide (CO), hydrocarbons (THC, Total Hydro Carbons), nitrogen oxides (NO_x), and particulate matter (PM, for direct-injection gasoline engines as from Euro 5) (Figures 7 and 8).

The limits for hydrocarbons and nitrogen oxides for the Euro 1 and Euro 2 stages are combined into an aggregate value ($HC+NO_x$). Since Euro 3 a special NO_x limit in addition to the aggregate value has applied to vehicles with diesel engines; in the case of gasoline vehicles, the aggregate has been replaced by separate HC and NO_x limits.

The Euro 5 stage was introduced in two steps as Euro 5a and Euro 5b. Euro 5a (from September 2009) saw the addition for gasoline engines of an NMHC limit (non-methane hydrocarbons), Euro 5b (from September 2011) for diesel engines a particulate-number limit of $6 \cdot 10^{11}$ particulates per kilometer. This PN limit has also applied from Euro 6b (September 2014) and Euro 6c (September 2017) to vehicles with direct-injection gasoline engines (a higher interim value of $6 \cdot 10^{12}$ particulates/km for Euro 6b is possible on request).

The limits are different for vehicles with diesel and gasoline engines, but with Euro 6 will be brought further into line.

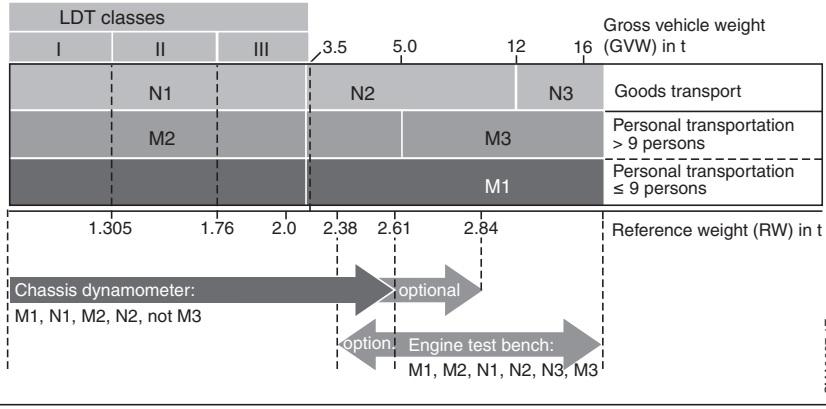
The limits for LDT class 1 are the same as for cars. Passenger cars with a permissible total in excess of 2.5 t were for Euro 3 and Euro 4 treated as LDTs and were therefore likewise categorized into one of the three LDT classes. This option ceases to apply from Euro 5.

The limits are defined based on mileage and indicated in grams per kilometer (g/km). Since Euro 3, emissions are measured on a chassis dynamometer using the MNEDC (Modified New European Driving Cycle).

Figure 6: Vehicle classes, EU legislation

LDT Light Duty Truck, MDV Medium Duty Vehicle, HDV Heavy Duty Vehicle, PC Passenger Car, LDV Light Duty Vehicle, LVW Loaded Vehicle Weight, GVW Gross Vehicle Weight.

Note: The axes for GVW and RW must be considered separately!



Since September 2017 the MNEDC has been replaced by the WLTC (Worldwide Light-duty Test Cycles, see also Test cycles). The dynamic test cycle derived from road driving is supplemented by a strongly revised test procedure (WLTP, Worldwide harmonized Light vehicles Test Procedure) which is to make possible significantly more realistic consumption and emission measurements. Here the Euro 6c limits defined for the MNEDC have not been adapted to the WLTC. This stage is called Euro 6d-temp since, aside from the change to the WLTC, the new "Real Driving Emissions" test (RDE) will also be introduced with a first stage. Euro 6d with the second RDE stage will apply from January 2020.

Evaporative emissions
The subject of evaporative emissions is addressed in the chapter "Exhaust-gas measuring techniques".

Type approval
Type approval is conducted along similar lines to that in the USA, with the following differences: Test vehicles absolve an initial run-in period of 3,000 kilometers before testing. The limit values in the Type I

Figure 7: EU emission limits for gasoline-engine passenger cars and light commercial vehicles [2]

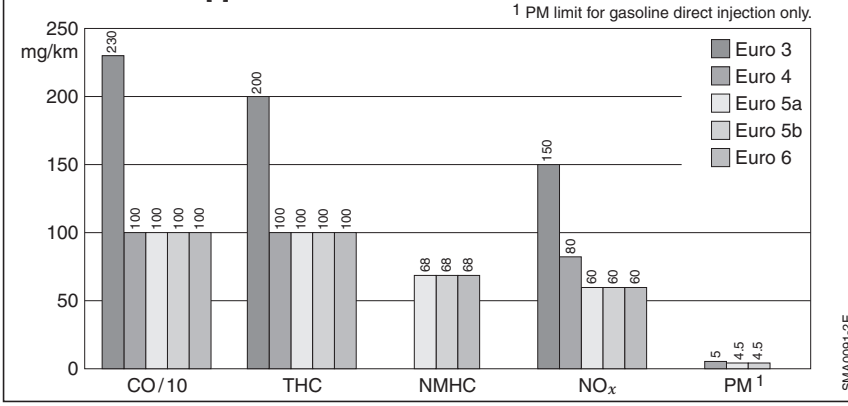
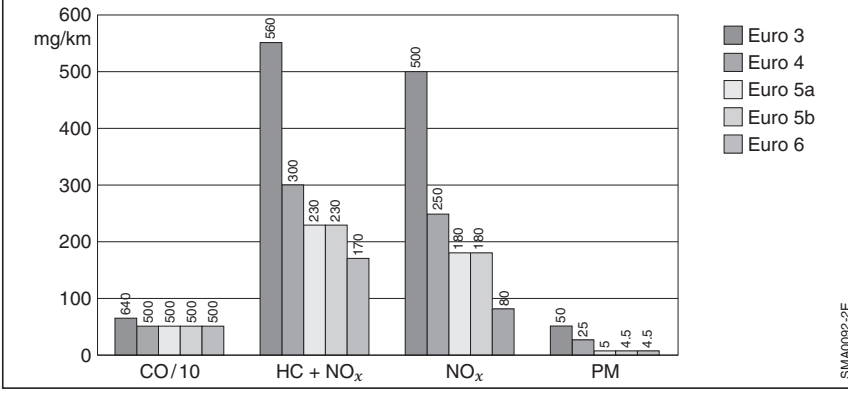


Figure 8: EU emission limits for diesel-engine passenger cars and light commercial vehicles [2]



test are "full useful life" limit values, i.e., they must be complied with even when the durability distance is reached. In order to take into account the aging of components up to the durability distance, deterioration factors are applied to the values measured in the type approval. These are defined in the legislation for every pollutant component; manufacturers are also allowed to present documentation confirming lower factors obtained during specified endurance testing (Type V test) over 80,000 km. From Euro 5 the endurance distance has been increased from 80,000 km to 160,000 km, where further alternative test procedures are possible.

Type tests

There are four essential type tests for type approval. Type I, Type IV, Type V and Type VI tests are used for vehicles with gasoline engines; only Type I and Type V tests are used for vehicles with diesel engines. With Euro 6d-temp the RDE test Type Ia and the Ambient Temperature Correction Test are added for gasoline and diesel (ATCT, Type I test at 14 °C for converting the CO_2 emission from 23 °C to 14 °C, emission limits must be adhered to).

With the Type I test, the primary exhaust-gas test, the exhaust emissions are determined after cold starting in the MNEDC (Modified New European Driving Cycle) and since Euro 6d-temp in the WLTC (Worldwide Light-duty Test Cycles). In addition, the opacity of the exhaust gas is recorded for vehicles with diesel engines.

The Type Ia test "Real Driving Emissions" (RDE) is intended to ensure that the emission limits are adhered to not only in the standardized cycle but also under real road conditions. In the EU the RDE test is used both during the type approval and also in in-field monitoring. To this end a vehicle is equipped with a mobile measurement system (PEMS, Portable Emissions Measurement System, see Exhaust-gas measuring techniques) and driven for 90 to 120 minutes in normal road traffic. Various boundary conditions apply during this test (see Table 3, Test cycles). CO (monitoring, not a limit value), NO_x (for gasoline and diesel), PN (particulate number, only for gasoline direct-injection engines and diesel), and CO_2 (as a normalization quantity) are measured. Valid RDE journeys are evaluated and weighted with software tools. The average NO_x and PN emissions in the urban part and for the entire journey are compared with maximum values which are obtained from the product of emission limit multiplied by "Conformity Factor" CF. For NO_x CF will be introduced in two stages (2.1 in 2017 and 1.43 in 2020), for PN 1.5 will apply from 2017. The CF values are to be adapted to the state-of-the-art (as 1 + 0.x).

Type IV testing measures evaporative emissions from parked vehicles. This primarily concerns fuel vapor that evaporates from the fuel-carrying system (fuel tank, pipes, etc.) (see also Exhaust-gas measuring techniques, also with regard to emission limit and test procedure).

Type VI testing embraces hydrocarbon and carbon-monoxide emissions immediately following cold starts at –7 °C. Only the first section (urban section) of the MNEDC is driven for this test. This test has been binding since 2002. A new test based on WLTC is being elaborated at UN ECE level.

The Type V test assesses the long-term durability of the emissions-reducing equipment. In addition to the specified endurance test, alternative test procedures are possible from Euro 5 (e.g., test-bench aging).

CO_2 emissions

For CO_2 emissions there were no limits laid down by law until 2011; however, the vehicle manufacturers did enter into a voluntary commitment in Europe. Because the manufacturers failed to reach their target, a fleet target value has been laid down by law for passenger cars. Within an introductory phase from 2012 through 2015 a fleet value of 130 g/km must be achieved (equating to 5.3 *l*/100 km gasoline or 4.9 *l*/100 km diesel). For light commercial vehicles a similar regulation stipulates a target value of 175 g/km for 2017. A further reduction for passenger cars to 95 g/km and for light commercial vehicles to 147 g/km has been decided on for 2020/2021. These values are to

understood as mean values for all the passenger cars and LDTs supplied by a vehicle manufacturer (vehicle fleet). The target value for a vehicle is dependent on the vehicle tare weight.

Since September 2017 new vehicle types have had to be certified in the WLTP (with four phases). The determined CO_2 values at 23 °C are corrected to 14 °C (by means of a factor determined in the Ambient Temperature Correction Test) and then converted with the aid of the CO2MPAS tool to MNEDC values, which were used to check whether the targets were achieved up to and including 2020. On the basis of the relative achievement of the target values in 2020 and the underlying MNEDC and WLTP values a manufacturer-specific target value will be defined for 2021 and subsequent years. Further reductions have been agreed for 2025 and 2030, but could be lowered further within the framework of the "Green Deal".

Similarly to the US CAFE regulations, fines must be paid if the target value is not met, however there is no option of credit/debit balancing over a number of years.

In-field monitoring

EU legislation also calls for conformity-verification testing on in-use vehicles as part of the Type I test cycle. The minimum number of vehicles of a vehicle type under test is three, while the maximum number varies according to the test procedure. Vehicles under test must meet the following criteria:

- Mileages vary from 15,000 km to 100,000 km, and vehicle age from 6 months to 5 years (from Euro 4).
- Regular service inspections were carried out as specified by the manufacturer.
- The vehicle must show no indications of non-standard use (e.g., tampering, major repairs, etc.).

If emissions from an individual vehicle fail substantially to comply with the standards, the source of the high emissions must be determined. If several vehicles display excessive emissions in random testing for the same reason, the results of the random test must be classified as negative. If there are various reasons, the test schedule may be extended, providing the maximum sample size is not reached.

If the type-approval authorities detect that a vehicle type fails to meet the criteria, the vehicle manufacturer must devise suitable action to eliminate the defect. The action catalog must be applicable to all vehicles with the same defect. If necessary, a recall action must be started.

Periodic emission testing

In Germany all passenger cars and light commercial vehicles and vans are required to undergo emissions inspections three years after their initial registration, and then at subsequent intervals of two years. For vehicles with gasoline engines, the main focus is on CO levels and λ closed-loop control, while for vehicles with diesel engines, the opacity test is the main criterion. Data from the diagnostic system are taken into consideration for vehicles with On-Board Diagnostic systems (OBD).

Comparable tests are also available in other countries; in Europe, for example, in Austria, France, Spain, and Switzerland, and in many parts of the USA in the form of "Inspection and Maintenance".

Japanese legislation

The permitted emission values are also subject to gradual stages of severity in Japan. The limit values have since September 2007 been subject to further tightening within the framework of the "New Long Term Standards".

Vehicles with diesel engines have since September 2010 been subject to further tightening of the limit values in the form of the "Post New Long Term Regulation". In the case of vehicles with gasoline engines, the previous synthetic test cycles were replaced in two stages (2008 and 2011) by the more realistic JC08 cycle.

In 2018 Japan also introduced the WLTC based on UN GTR No. 15 [3]. Unlike the EU, only the first three phases apply. The NMHC limit values for vehicles with gasoline engines and the NO_x limit values for vehicles with diesel engines are adapted.

Vehicle classes
Vehicles with a permissible total weight up to 3.5t are essentially divided into three classes (Figure 9): passenger cars (up to ten seats), LDVs (Light-Duty Vehicles) up to 1.7t, and MDVs (Medium-Duty Vehicles) up to 3.5t. MDVs have higher limits for CO and NO_x emissions (for vehicles with gasoline engines) than the other two vehicle classes. For diesel engines, the vehicle categories are distinguished by NO_x and particulate limits.

Emission limits
The Japanese legislation specifies limits for carbon monoxide (CO), nitrogen ox-

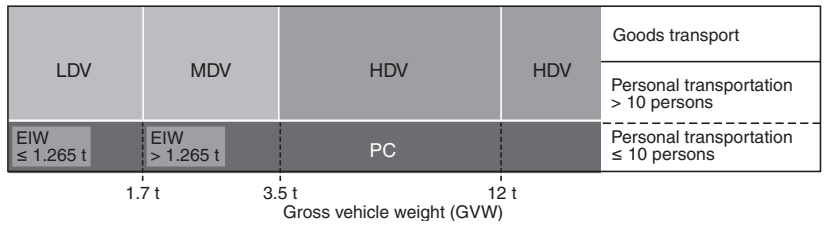
Table 1: Emission limits of Japanese legislation for gasoline-engine passenger cars

Vehicle category	Pass. cars, mini delivery vans, small commercial vehicles GVW ≤ 1.7 t		Medium-duty commercial vehicles 1.7 t < GVW 3.5 t	
Cycle Year (new types)	JC08 10/2009	WLTC 10/2018	JC08 10/2009	WLTC 10/2018
Limit values for gasoline engines				
NMHC [mg/km]	50	100	50	150
NO_x [mg/km]	50	50	70	70
CO [mg/km]	1,150	1,150	2,550	2,550
PM [mg/km] (direct-injection engines only)	5	5	7	7

Table 2: Emission limits of Japanese legislation for diesel-engine passenger cars

Vehicle category	Pass. cars, mini delivery vans, small commercial vehicles GVW ≤ 1.7 t		Medium-duty commercial vehicles 1.7 t < GVW 3.5 t	
Cycle Year (new types)	JC08 10/2009	WLTC 10/2018	JC08 10/2009	WLTC 10/2018
Limit values for diesel engines				
NMHC [mg/km]	24	24	24	24
NO_x [mg/km]	80	150	150	240
CO [mg/km]	630	630	630	630
PM [mg/km]	5	5	7	7

Figure 9: Vehicle classes, Japanese legislation
MDV Medium Duty Vehicle, HDV Heavy Duty Vehicle, PC Passenger Car, LDV Light Duty Vehicle, GVW Gross Vehicle Weight, EIW Equivalent Inertia Weight.



ides (NO_x), non-methane hydrocarbons (NMHC), particulate matter (for vehicles with diesel engines, from 2009 also for gasoline direct injection with leaner NO_x-reducing technology) (Table 1 for gasoline engines and Table 2 for diesel engines).

Pollutant emissions were determined with a combination of 11-mode and 10•15-mode test cycles (see Japanese test cycle). Cold-start emissions were thus also taken into account. A new test cycle was introduced in 2008 (JC08). This replaced initially the 11-mode and as from 2011 also the 10•15-mode, so that only the JC08 will be used as the cold- and hot-start test. Since 2018 the JC08 has been replaced by the WLTC.

Evaporative emissions
The exhaust-gas regulations in Japan include limits on evaporative emissions in vehicles with gasoline engines, which are measured using the SHED method (see Exhaust-gas measuring techniques). Japan is adopting the WLTC-based test procedure of UN GTR 19 [4], which has been influenced by the developments of the EU test procedure. Different tank systems are taken into account.

Durability
Manufacturers must demonstrate a durability of 45,000 km (New Long Term Standards) or 80,000 km (Post New Long Term Standards) for vehicles with diesel engines. 80,000 km apply to all stages for vehicles with gasoline engines.

Fleet consumption
In Japan targets apply to the fleet consumption of a manufacturer for 2010 and 2015, based on target values for vehicle weight classes. For tax-incentive purposes (green tax program) there are two stages rewarding fuel consumption which is improved by 15% and 25% respectively.

Chinese legislation
China has, up to and including Stage 5 (applicable nationally from 2018), adopted the EU legislation to limit exhaust-gas emissions. Leading the way here was Beijing, introducing the stages early, followed by further metropolitan areas such as Shanghai and Guangzhou. With China 6a (7/2020) and 6b (7/2023) China is introducing its own legislation which for exhaust-gas emissions links the WLTC with Euro 6-based and US limit values and adds its own requirements.

For the Type I test the 4-phase WLTC including WLTP of UN GTR 15 [3] is adopted. The limit values are fuel-neutral. For passenger cars and light commercial vehicles (class 1) for China 6a the Euro 6 limit values for gasoline engines apply for HC, NMHC, and NO_x. 0.7 g/km has been defined for the CO limit value. For PM the limit value 4.5 mg/km applies, for PN $6 \cdot 10^{11}$ particulates/km (no exception for gasoline engines with intake-manifold injection). A limit value for N_2O (laughing gas, a greenhouse gas) of 20 mg/km has been adopted from the USA. For China 6b the limit values for HC, NMHC, CO, and NO_x are reduced to approximately 50% of the Euro 6 limit values for gasoline engines. PM is reduced to 3 mg/km, PN and N_2O remain the same.

The durability requirement is 160,000 km for China 6a and 200,000 km for China 6b.

The Type VI test (at –7 °C) applies to vehicles with gasoline and diesel engines as of China 6a. In the first two phases of the WLTC limit values must be observed for HC of 1.2 g/km, CO of 10 g/m, and NO_x of 0.25 g/km.

China is also introducing an RDE test for the type approval and in-field monitoring. As of China 6a RDE tests must be conducted ("monitoring" of CO, NO_x, and PN, without limit values), as of China 6b "Conformity Factors" CF apply of 2.1 for NO_x and PN for all gasoline and diesel engines. The test procedures are based on an intermediate stage of the EU RDE legislation, e.g., the cold-start period is not taken into consideration and additional boundary conditions are defined, e.g., measurements up to an altitude of 2,400 m and lower top speeds for freeway driving. A review of the CF is envisaged by mid-2022.

The metropolitan areas could introduce China 6a and can introduce China 6b early, but restrictions apply, e.g., before

7/2023 for RDE only monitoring and 160,000 km as the durability distance.

China has also developed its own requirements for limiting fuel consumption and thus CO_2 emissions. In four stages (2005 through 2007, 2008 through 2015, 2016 through 2020, and 2021 through 2025) limit values dependent on the vehicle weight in liters of fuel per 100 km apply to individual vehicle types. In parallel there is a Corporate Average Fuel Consumption regulation (CAFC), which similarly to the EU CO_2 fleet targets sets target values dependent on vehicle weight for the consumption for the fleet of a manufacturer in the periods 2012 through 2015 and 2016 through 2020. Target values for the entire fleet are 6.9 *l* gasoline per 100 km for 2015 and 5.0 *l* gasoline per 100 km for 2020. For 2025 the target value is 4.0 *l* gasoline per 100 km.

The subject of evaporative emissions (Type IV test) is addressed in the chapter "Exhaust-gas measuring techniques".

China uses a variety of measures, such as for example tax reductions, to promote the introduction of electrified vehicles (hybrids and plug-in hybrids, pure electric and fuel-cell vehicles). With the "New Energy Vehicle" legislation (NEV) manufacturers must from 2019 fulfill quotas for the sale of these vehicles (cf. CARB ZEV legislation).

Test cycles for passenger cars and light commercial vehicles

US test cycles

FTP 75 test cycle

The FTP 75 test cycle (Federal Test Procedure, Figure 10a) consists of speed cycles that were actually recorded in commuter traffic in Los Angeles. This test is also in force in some countries of South America and in Korea besides the USA (including California).

Conditioning

The vehicle is subjected to an ambient temperature of 20 to 30 °C for a period of 6 to 36 hours.

Collection of pollutants

The vehicle is started and driven on the specific speed cycle on a chassis dynamometer. The emitted pollutants are collected in separate bags during defined phases (see Exhaust-gas measuring techniques).

Phase ct (cold transient)

The exhaust gas is collected during the cold test phase (0 to 505 s).

Phase cs (cold stabilized)

The stabilized phase begins 506 seconds after start. The exhaust gas is collected without interrupting the driving cycle. Upon termination of phase cs, after a total of 1,372 seconds, the engine is switched off for a period of 600 seconds (hot soak).

Phase ht (hot transient)

The engine is restarted for the hot test. The speed cycle is identical to the cold transient phase (Phase ct).

Phase hs (hot stabilized)

For hybrid vehicles, a further phase hs is driven. It corresponds to the progression of phase cs. For other vehicles, it is assumed that the emission values are identical to the cs phase.

Assessment
The bag samples from the first two phases are analyzed during the pause before the hot test. This is because samples may not remain in the bags for longer than 20 minutes.

The sample exhaust gases contained in the third bag are also analyzed on completion of the driving cycle. The total result includes emissions from the three phases rated at different weightings.

The pollutant masses of phases ct and cs are aggregated and assigned to the total distance of these two phases. The result is then weighted at a factor of 0.43.

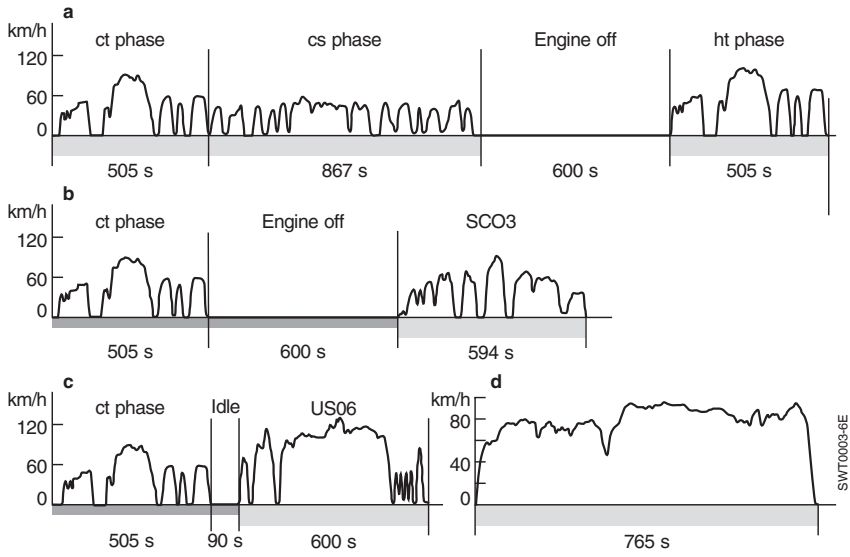
The same process is applied to the aggregated pollutant masses from phases ht and cs, related to the total distance of these two phases, and weighted at a factor of 0.57. The test result for the individual pollutants (among others HC, CO, and NO_x) is obtained from the sum of the two previous results.

The emissions are specified as the pollutant emission per mile.

SFTP cycles
The tests in accordance with the SFTP standard (Supplemental Federal Test Procedure) were introduced in 2001. These are composed of two driving cycles, the SC03 cycle (Figure 10b) and the US06

Figure 10: US test cycles for passenger cars and light commercial vehicles

	a	b	c	d
Test cycle	FTP 75	SC03	US06	Highway
Cycle distance:	17.87 km	5.76 km	12.87 km	16.44 km
Cycle duration:	1,877 s + 600 s pause	594 s	600 s	765 s
Mean cycle speed:	34.1 km/h	34.9 km/h	77.3 km/h	77.4 km/h
Maximum cycle speed:	91.2 km/h	88.2 km/h	129.2 km/h	96.4 km/h



cycle (Figure 10c). The extended tests are intended to examine the following additional driving conditions:
– Aggressive driving,
– Radical changes in vehicle speed,
– Engine start and acceleration from a standing start,
– Operation with frequent minor variations in speed,
– Periods with vehicle parked,
– Operation with air conditioner on.

For preconditioning, the SC03 and US06 cycles proceed through the ct phase from FTP 75 without exhaust-gas collection. However, other conditioning procedures are also possible.

The SC03 cycle (for vehicles with air conditioning only) is carried out at a temperature of 35 °C and 40 % relative humidity. The individual driving schedules are weighted as follows:
– Vehicles with A/C systems:
  35 % FTP 75 + 37 % SC03
  + 28 % US06.
– Vehicles without A/C systems:
  72 % FTP 75 + 28 % US06.

The SFTP and FTP 75 test cycles must be successfully completed on an individual basis.

Cold-start enrichment, which is necessary when a vehicle with a gasoline engine is started at low temperatures, produces higher emissions. These cannot be measured in current emissions testing, which is conducted at ambient temperatures of 20 to 30 °C. An additional exhaust-gas test is performed at –7 °C on vehicles with gasoline engines in order to limit these pollutants. However, this test only prescribes a limit for carbon monoxide; a nationally applicable fleet limit was introduced for NMHC emissions in 2013.

Test cycles for determining fleet consumption
Each vehicle manufacturer is required to provide data on corporate average fuel economy. Manufacturers that fail to comply with the target values are required to pay penalties.

Fuel consumption is determined from the exhaust-gas emissions produced during two test cycles – the FTP 75 test cycle (weighted at 55 %) and the highway test cycle (weighted at 45 %). An unmeasured highway test cycle (Figure 10d) is conducted once after preconditioning (vehicle allowed to stand with engine off for 12 hours at 20 to 30 °C). The exhaust-gas emissions from a second test run are then collected. The CO_2 emissions are used to calculate fuel consumption.

Further test cycles
FTP 72 test cycle
The FTP 72 test routine – also known as the UDDS (Urban Dynamometer Driving Schedule) – corresponds to the FTP 75 test, but does not include the ht test component (hot test). This cycle is driven during the running-loss test for vehicles with a gasoline engine.

New York City Cycle (NYCC)
This cycle is also an element in the running-loss test (for vehicles with a gasoline engine). It simulates low speeds in urban traffic with frequent stops.

Hybrid cycle
For hybrid vehicles the phase hs (progression corresponding to phase cs) is attached to the FTP 75 cycle. This driving cycle thus corresponds twice to the UDDS cycle, which is why it is called 2UDDS.

European test cycle
MNEDC
The “Modified New European Driving Cycle” (MNEDC, Figure 11) has been in force since Euro 3. Contrary to the “New European Driving Cycle” (Euro 2), in which measurement of emissions only began 40 seconds after the vehicle was started, the MNEDC also includes a cold-start phase (including engine starting).

Conditioning
The vehicle is subjected to an ambient temperature of 20 to 30 °C for a minimum period of six hours. Since 2002 the starting temperature has been lowered to –7 °C for the Type VI test (only for vehicles with gasoline engines).

Collection of pollutants
The exhaust gas is collected in bags in two phases: the Urban Driving Cycle (UDC) at a top speed of 50 km/h and the Extra Urban Driving Cycle (EUDC, overland drive) at a top speed of 120 km/h.

Assessment
The pollutant mass measured by analyzing the bag contents is referred to the distance covered (see Exhaust-gas measuring techniques).

WLTC
Within the framework of the UN ECE a new test cycle WLTC (Worldwide Light-duty Test Cycles) was devised by the EU, Japan, India, and South Korea (among others) for passenger cars and light commercial vehicles and defined in UN GTR No. 15 [3]. The WLTC consists of four phases ("low", "mid", "high", and "extra high speed"). As an option, however, only the first three phases can be used. Further cycles were developed for special vehicle segments: for Japanese "kei cars" a toned-down variant of the WLTC and for the Indian market two cycles for vehicles with very low power-weight ratios ("low powered vehicle test cycles", LPTCs).

The dynamic test cycle derived from road driving is supplemented by a strongly revised test procedure (WLTP) compared with the previous regulation which is to make possible significantly more realistic consumption and emission measurements.

Among other things, the following have been defined for the WLTP test procedure: a method for shift-point calculation for manual transmissions, higher vehicle weights by taking into account payload and equipment options, improved methods for determining the coasting parameters, stepless adjustment of the chassis dynamometer, and 23 °C as the test temperature. Aside from the fundamental test procedure, special requirements for electrified vehicles such as hybrid and electric vehicles have also been elaborated.

Figure 12 shows the WLTC by way of example for class 3b, the class with the biggest-selling passenger cars and light commercial vehicles. Compared with the previous test cycle MNEDC (Figure 11),

Figure 11: MNEDC for passenger cars and light commercial vehicles
Cycle distance: 11 km.
Average speed: 33.6 km/h.
Top speed: 120 km/h.
UDC Urban Driving Cycle,
EUDC Extra Urban Driving Cycle.

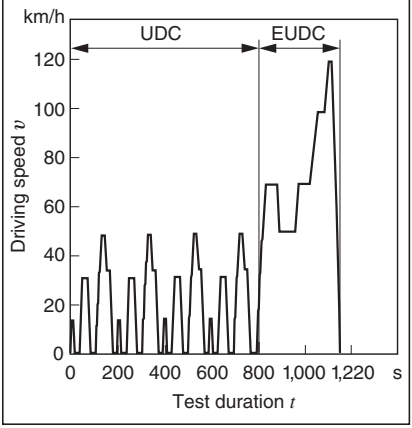
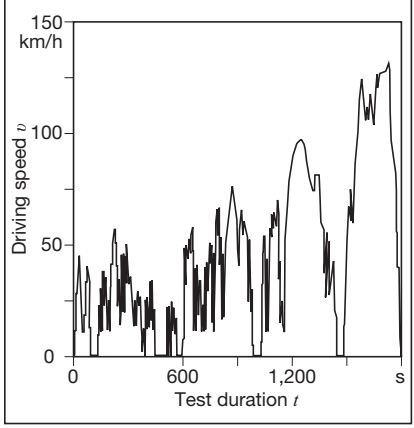


Figure 12: WLTC for passenger cars and light commercial vehicles



the changed dynamics of the WLTC can be clearly discerned.

RDE test

A new element of the legislation for passenger cars and light commercial vehicles is the RDE test ("real driving emissions"), which was developed in the EU and is being adopted by other countries like China and India (for gasoline and diesel) and Korea and Japan (for diesel only). The aim of the RDE test is to ensure that the emissions limits are observed not only in the standardized cycle but also under real road conditions. For the RDE test a vehicle is equipped with a mobile measurement system (PEMS, Portable Emissions Measurement System, see Exhaust-gas measuring techniques) and driven in normal road traffic. Various boundary conditions apply during this test, e.g., with regard to the proportions of city and intercity roads and freeway, the proportions of idle, of average speed, of minimum and maximum speeds, acceleration, cold start, and ambient conditions such as temperature and altitude. Table 3 sets out selected boundary conditions for the EU. RDE journeys can be made under conditions that are quite different from those in the Type I test (23 °C) thanks to the enhanced temperatures of –7 °C to +35 °C.

PEMS systems are available for CO, NO_x, PN (particulate number), and CO_2. Valid RDE journeys (see Table 3) are evaluated and if necessary weighted with software tools. The average emissions for a PEMS journey are compared with maximum values which are obtained from the product of emission limit multiplied by "Conformity Factor" CF. The CF takes into account the measuring inaccuracies

Table 3: Boundary conditions for a valid RDE measurement run (selection, not all specified)

Parameter	Criterion		
Route distribution	Urban 34 % (29 to 44 %)	Interurban 33 % (±10 %)	Freeway 33 % (±10 %)
	In each case at least 16 km		
Speeds	$v \leq 60$ km/h v_D = 15 to 40 km/h (average speed)	60 km/h < $v \leq 90$ km/h	90 km/h < $v \leq 145$ km/h (up to 160 km/h for max. 3 % of the freeway time)
Total driving time	90 to 120 min		
Secondary loads	Operation of the air-conditioning system and the other secondary loads must correspond to their possible use by the load under normal driving conditions on the road.		
Altimeter	Moderated: $h \leq 700$ m Extended: 700 m < $h \leq 1{,}300$ m (measured values are divided by 1.6) Starting and end points may not differ in their height above sea level by more than 100 m.		
Temperature	Moderated: Stage 1: 3 °C ≤ $T \leq 30$ °C Stage 2: 0 °C ≤ $T \leq 30$ °C Extended (measured values are divided by 1.6): Stage 1: –2 °C ≤ T < 3 °C, or 30 °C < $T \leq 35$ °C Stage 2: –7 °C ≤ T < 0 °C, or 30 °C < $T \leq 35$ °C		
Vehicle capacity weight and test weight	Basic capacity weight: driver, front passenger, test equipment. Maximum weight from basic capacity weight and artificial capacity weight: max. 90 % of the sum total of the "weight of passengers" and the "capacity weight".		
Cold start	Cold-start period: until the coolant has reached 70 °C, max. 5 minutes cumulative combustion-engine operation. Speed: average = 15 to 40 km/h, max. 60 km/h. Emissions in the cold-start period are evaluated for the urban part and the total RDE journey with the normal methods.		

which arise from measurements with PEMS conducted on the road compared with a measurement conducted on a chassis dynamometer.

Japanese test cycle

JC08 test cycle

In 2008 a new exhaust-gas test was introduced in the form of the JC08 (Figure 13), which initially replaced the 11-mode test as the cold-start test. Since 2011 only the JC08 has continued to be used (both as the cold-start and the hot-start tests). The cold-start test is weighted at 25 %, the hot-start test at 75 %. The pollutants are converted based on distance traveled, i.e., into grams per kilometer (g/km).

WLTC

Since 2018 the WLTC with three phases (WLTC without the "extra high speed" part) has replaced the JC08.

Emission-control legislation for heavy commercial vehicles

US legislation

Vehicle classes

Heavy commercial vehicles are defined in EPA legislation as vehicles with a gross vehicle weight over 8,500 lbs or 10,000 lbs (equivalent to 3.9 t or 4.6 t), depending on vehicle type (see Figure 4).

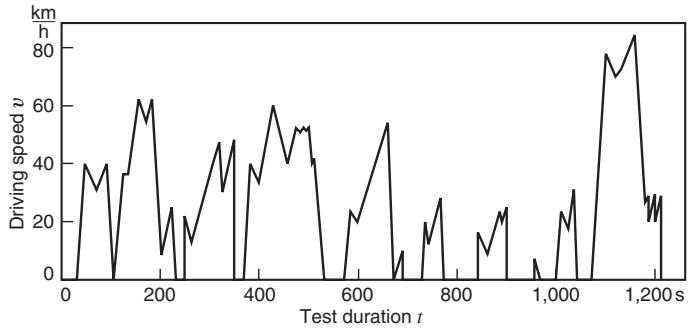
In California, all vehicles over 14,000 lbs (6.35 t) are classified as heavy commercial vehicles (see Figure 1). To a great extent, Californian legislation is identical to parts of EPA legislation. However, there is an additional program for city buses.

Emission limits

The US standards specify limits for diesel engines for hydrocarbons (HC), carbon monoxide (CO), nitrogen oxides (NO_x), particulate matter (PM), exhaust-gas opacity, and in parts for non-methane hydrocarbons (NMHC).

The permissible limits are related to engine power output and specified in g/bhp-h (gram per brake horsepower hour) (Figure 14 with values converted to g/kWh). The emissions are measured on the engine test bench during the dynamic test cycle with cold-starting sequence

Figure 13: Japanese test cycle JC08 for passenger cars and light commercial vehicles
Cycle distance: 8.179 km.
Cycle time: 1,204 s.
Average speed: 24.5 km/h.
Top speed: 81.6 km/h.



(HDDTC, Heavy-Duty Diesel Transient Cycle); the exhaust-gas opacity is measured using the Federal Smoke Test (FST).

New, more stringent regulations have applied to vehicles since model year 2004, with significantly reduced NO_x emission limits. Non-methane hydrocarbons and nitrogen oxides are grouped together in one aggregate (NMHC + NO_x).

Another very drastic tightening of emission restrictions came into force in model year 2007. The NO_x and particulate emissions are separately limited and are a tenth of the previous values. These are not achievable without the use of exhaust-gas treatment systems (e.g., NO_x reduction measures with NO_x storage catalytic converter or active SCR system, particulate filter).

A gradual phase-in took place for NO_x and NMHC emission limits between model years 2007 and 2010.

From model year 2024 with the introduction of the CARB Low NO_x Regulation (for buses) the nitrogen-oxide limits will be reduced by 90 % and the previous particulate-emission limit halved in California.

Massively increased durability requirements will additionally apply from model year 2027.

A binding share of sales of zero-emission vehicles is prescribed in California for model years 2024 through 2035 (Advanced Clean Truck Regulation, ACT).

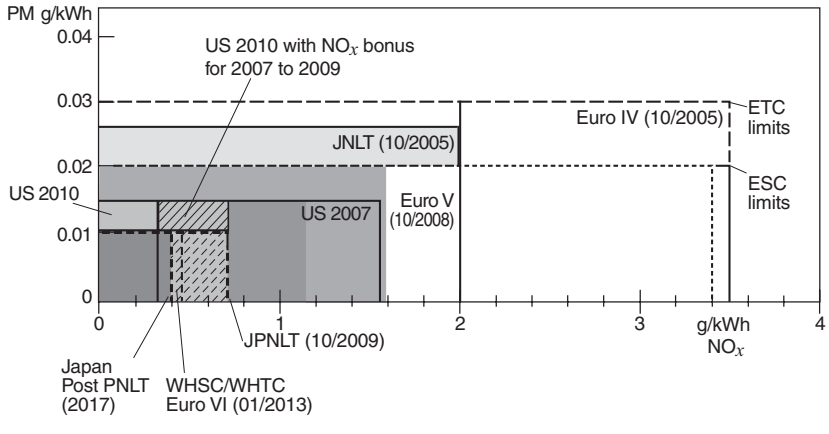
To help compliance with severe particulate limits, the maximum permitted sulfur content in diesel fuel was reduced to 15 ppm from mid-2006.

For heavy commercial vehicles – in contrast with cars and light commercial vehicles – there are no limits specified for average fleet emissions and fleet consumption.

Consent Decree

In 1998 a legal agreement was reached between EPA, CARB, and a number of engine manufacturers. It provides for sanctions against manufacturers if they make illegal modifications to engines to achieve optimized consumption in the highway cycle, resulting in higher NO_x emissions. The “Consent Decree” specifies that the applicable emission limits must also undercut the steady-state European 13-stage test in addition to the dynamic test cycle. Furthermore, emissions are not allowed to exceed the limits for model year 2004 by more than 25 %, regardless of driving mode within a specified engine-speed/torque range (Not-to-Exceed Zone).

Figure 14: Comparison of NO_x and PM emission limits for diesel fuel commercial vehicles: EU, USA, Japan [5]



These additional tests have been mandatory for all diesel commercial vehicles since model year 2007. However, emissions in the not-to-exceed zone may be up to 50 % above the emission limits.

Durability
Compliance with emission limits must be demonstrated over a defined mileage or a specific time period in which a distinction is drawn between three classes, each with increasing durability requirements:
– Light commercial vehicles of 8,500 lbs (EPA) and 14,000 lbs (CARB) to 19,500 lbs: 10 years or 110,000 miles.
– Medium-duty commercial vehicles from 19,500 lbs to 33,000 lbs: 10 years or 185,000 miles.
– Heavy commercial vehicles over 33,000 lbs: 10 years or 435,000 miles and 22,000 h (operating time).

Fuel-consumption requirement
In the USA there are separate requirements with regard to greenhouse-gas emissions (since 2014) and fuel consumption (since 2017).

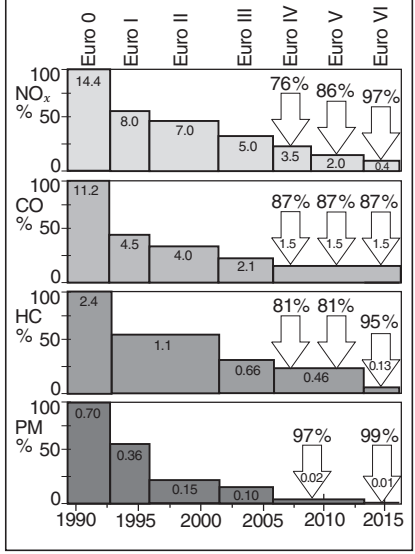
EU legislation
Vehicle classes
In Europe, all vehicles with a permissible gross vehicle weight of over 3.5 t, or capable of transporting more than nine persons, are classified as heavy commercial vehicles (see Figure 6). The emission regulations (Euro standards) are based on the Basic Directive 88/77/EEC [6], which is continuously updated and extended. Euro VI requirements are described by the created Directive (EU) 64/2012 [7]. Supplements to PEMS (Portable Emission Measurement System) and IUC (In Use Compliance) can be found for example in Regulation 2016/1718 EU [8]. The most up-to-date supplement to Euro VI can be found in Regulation (EU) 2020/1181 [9].

Emission limits
As for passenger cars and light commercial vehicles, new emission-limit levels for heavy commercial vehicles are introduced in two stages. New engine designs must meet the new emission limits during type approval. One year later compliance with the new emission limits is a prerequisite for registering a new vehicle. The legislator can inspect conformity of production (COP) by taking engines out of serial production and testing them for compliance with the new emission limits.

For commercial-vehicle diesel engines, the Euro standards define emission limits for hydrocarbons (HC and NMHC), carbon monoxide (CO), nitrogen oxides (NO_x), particulates, and exhaust-gas opacity. The permissible limits are related to engine power output and specified in g/kW (Figure 15).

In October 2000 the limit stage Euro III was binding initially on new engine types and one year later on all newly manufactured engines. Emissions were measured during the 13-stage European Steady-State Cycle (ESC, see Test cycles for heavy commercial vehicles), and exhaust-gas opacity in the supplementary European Load Response (ELR) test. Diesel engines which are equipped with systems for exhaust-gas treatment

Figure 15: EU emission limits for diesel commercial vehicles [5]
Values in g/kWh.
Percentage reductions refer to Euro 0.



(NO_x reduction measures with NO_x storage catalytic converter or active SCR system, particulate filter) furthermore already had to be tested in the dynamic exhaust-gas emission test ETC (European Transient Cycle). These European test cycles were conducted with the engine running at normal operating temperature. Since Euro V a cold-start test has also been required.

Within the Euro III legislation a distinction was still made between large engines (displacement > 0.75 l per cylinder) and small engines (displacement < 0.75 l per cylinder and rated speed > 3,000 rpm), with higher permissible particulate-emission limits for the smaller units. This distinction is dropped for new certifications with the introduction of Euro IV.

In October 2005 the Euro IV emission-limit level came into force initially for new type approvals, and for serial production one year later. All emission limits were significantly lower than specified by Euro III, but the biggest increase in severity applied to particulates, for which the limits were reduced by approximately 80 %. The following changes also applied after introduction of Euro IV:

– The dynamic exhaust-gas emission test (ETC) was obligatory – in addition to ESC and ELR – for all diesel engines.
– The continued functioning of emissions-related components must be documented for the entire service life of the vehicle (see Durability).

The Euro V emission-limit level was introduced in October 2008 for all new engine approvals, and one year later for all new serial-production vehicles. Only the NO_x emission limits were more severe compared to Euro IV.

In January 2013 the Euro VI emission-limit level for new engine types came into force (year later for all newly produced engines). Compared with Euro V, the nitrogen-oxide emissions are again reduced by 80 % and the particulate emissions by more than 60 % (referred to ETC limits for Euro V). New harmonized engine tests have been introduced with Euro VI. Here, too, there is a stationary test (WHSC, World Harmonized Stationary Cycle), a dynamic test (WHTC, World Harmonized Transient Cycle), and random tests within specified engine-speed/torque ranges (WNTE, World Harmonized Not-to-Exceed Zone). Unlike the previous Euro V regulations, from Euro VI no specific particulate-matter limits will be given for the transient test; the limits given will be identical to those for the stationary test. Particulate-number limits are additionally demanded with Euro VI (separately for stationary and transient tests).

Durability

Compliance with emission limits must be demonstrated over a defined mileage or a specific time period in which a distinction is drawn between three classes, each with increasing durability requirements:

– Light commercial vehicles up to 3.5 t gross vehicle weight (GVW): 6 years or 100,000 km (Euro IV and Euro V) and 160,000 km (Euro VI).
– Medium-duty commercial vehicles smaller than 16 t GVW: 6 years or 200,000 km (Euro IV and Euro V) and 300,000 km (Euro VI).
– Heavy commercial vehicles over 16 t GVW: 7 years or 500,000 km (Euro IV and Euro V) and 700,000 km (Euro VI).

Fuel consumption

CO_2 emissions from heavy commercial vehicles with a permissible gross vehicle weight of more than 16 t are to be reduced by 15 % by 2025 and by 30 % by 2030. These targets signify a relative lowering referred to average determined emissions from vehicles of a manufacturer with a reference time period of 07/2019 through 06/2020.

Subcategories are created with regard to the CO_2 output values, depending on vehicle type (box body or semitrailer truck) and axle configuration.

CO_2 is calculated by means of a simulation tool (VECTO, Vehicle Energy and Consumption Calculation tool), which takes into account a large number of parameters (vehicle application, drivetrain configuration, etc.).

Very low-emission vehicles
Prior to the introduction of the Euro V standard it was possible to meet voluntarily at an early stage more demanding, so-called EEV limit values (Enhanced Environmentally-Friendly Vehicle). Voluntary early meeting of these limits permitted tax incentives. EEV limit values for HC, NMHC, CO, and exhaust-gas opacity were lower than the Euro V limit values. NO_x and particulate limits were equivalent to the Euro V ESC limits. Natural-gas-powered city and tourist buses are typical examples of EEV vehicles.

Japanese legislation
Vehicle classes
In Japan, vehicles with a permissible gross vehicle weight of over 3.5 t, or capable of transporting more than ten persons are classified as heavy commercial vehicles (see Figure 9).

Emission limits
In October 2005 the "New Long Term Regulation" was introduced and was valid until the end of 2009. It stipulated emission limits for hydrocarbons (HC), nitrogen oxides (NO_x), carbon monoxide (CO), particulates, and exhaust-gas opacity. Emission levels were measured in the newly introduced transient JE05 test cycle (hot test), and exhaust-gas opacity in the Japanese smoke test. The smoke test was conducted with three full-load operating points at 40.6 % and 100 % rated speed. With the introduction of the "Post New Long Term" emissions stage (09/09) there are now no smoke-test specifications.

The "Post New Long-Term Regulation" came into force in September 2009. The particulate and NO_x limits were reduced by almost two-thirds compared with 2005 levels.

Since October 2016 the "Post Post New Long Term" (Post PNLT, PPNLT) has applied to vehicles > 7.5 t with a renewed NO_x reduction to approximately 60 % of the previous PNLT requirements. The introduction of PPNLT will be completed in October 2018 also with limits for commercial vehicles ≤ 7.5 t. The deadlines at a glance:

– 10/2016 for heavy commercial vehicles > 7.5 t gross vehicle weight (GVW, not including semitrailer trucks),
– 10/2017 for heavy semitrailer trucks > 7.5 t GVW,
– 10/2018 for vehicles with 3.5 to 7.5 t GVW.

The harmonized test cycles WHSC and WHTC are used as the test cycle.

Durability
Compliance with the emission limits must be demonstrated over a defined mileage in which a distinction is drawn between three classes, each with increasing durability requirements.
– Commercial vehicles smaller than 8 t gross vehicle weight (GVW): 250,000 km.
– Medium-duty commercial vehicles smaller than 12 t GVW: 450,000 km.
– Heavy commercial vehicles over 12 t GVW: 650,000 km.

Fuel-consumption requirement
Fuel-consumption limits are prescribed for trucks and buses with a permissible gross vehicle weight of over 3.5 t. Two driving cycles (urban and extra-urban) are used.

Fuel consumption is determined on the engine test bench. Because consumption is very much dependent on the individual vehicle engine and equipment specification (e.g., drivetrain, rolling resistance, vehicle weight), the calculation is made using a conversion program. The requirements are for a commercial vehicle with a gross vehicle weight smaller than 10 t 13.4 l/100 km, for semitrailer trucks with a gross vehicle weight smaller than 20 t 32.1 l/100 km/l, and for buses with a gross vehicle weight smaller than 14 t 18.9 l/100 km. The figures apply to 2025.

Regional programs
In addition to the nationwide regulations for new vehicles, there are also regional requirements for the overall vehicle population aimed at reducing existing emission levels by replacing or upgrading old diesel vehicles.

The "Vehicle NO_x Law" has been in force since 2003 within, among other places, the greater urban area of Tokyo to vehicles with a gross vehicle weight of over 3,500 kg. It states that 8 to 12 years after a vehicle is first registered, the NO_x and particulate limits of the relevant preceding phase of emission limits must be adhered to. The same principle also applies to particulate emissions. Here, the regulation will already apply seven years after first vehicle registration.

Test cycles for heavy commercial vehicles

For heavy commercial vehicles, all test cycles are run on the engine test bench. In the transient test cycles, the emissions are collected and evaluated according to the CVS principle. The untreated emissions are measured in the stationary test cycles. Emissions are specified in g/kWh.

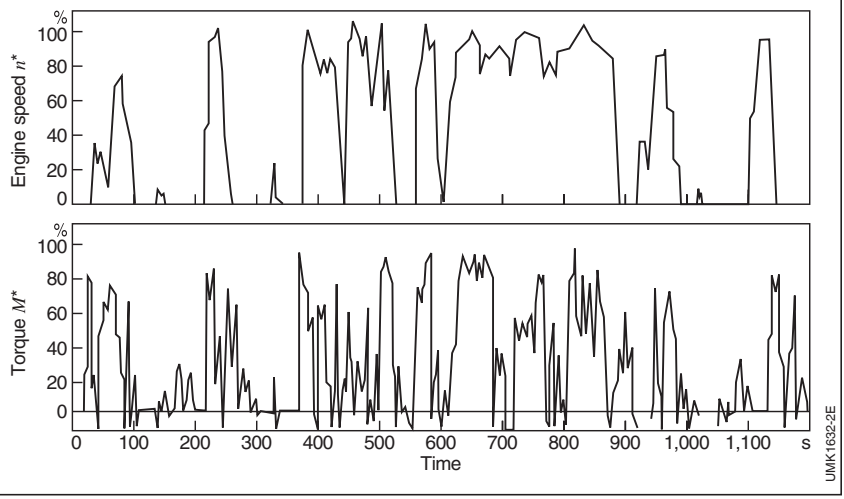
USA

EPA Engine Dynamometer Schedule for Heavy-Duty Diesel Engines (HD FTP Transient)

Since 1987 engines for heavy commercial vehicles have been tested from a cold start on an engine test bench in a transient driving cycle (US HDDTC, Heavy Duty Diesel Transient Cycle). The test cycle is basically equivalent to operating an engine under realistic road-traffic conditions (Figure 16). It includes significantly more idle sections than the WHTC (see section "Europe").

Figure 16: US FTP (Heavy-Duty Diesel Transient Cycle, HDDTC) for heavy commercial-vehicle engines

Both the nominal engine speed n^* and the nominal torque M^* are taken from tables specified by legislation.



Federal Smoke Cycle
An additional test, the Federal Smoke Cycle, tests exhaust-gas opacity under dynamic and quasi steady-state conditions. In this test, for different load conditions (torque specifications), a number of sudden full-load accelerations are run on an engine test bench to determine exhaust-gas opacity.

Additional test cycle
Since model year 2007 US emission limits have also had to comply with the European 13-stage test (ESC). Furthermore, emissions in the not-to-exceed zone (i.e., with any driving mode within a specified engine-speed/torque range) may be max. 50 % above the emission limits.

Figure 17: Steady-state 13-stage test ESC (Europe)

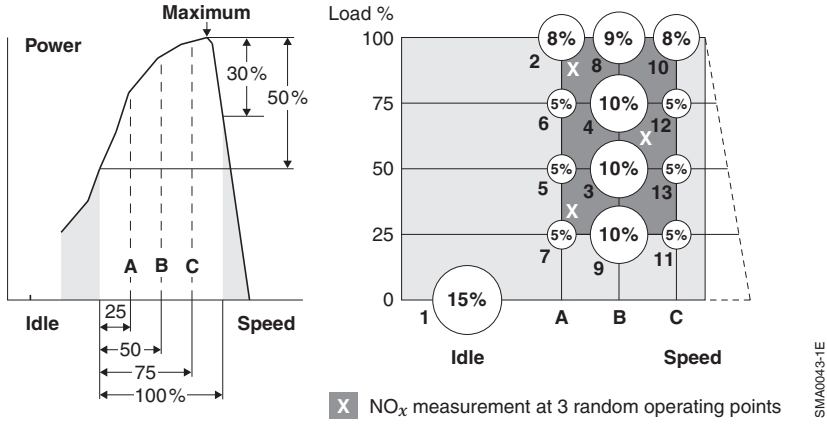
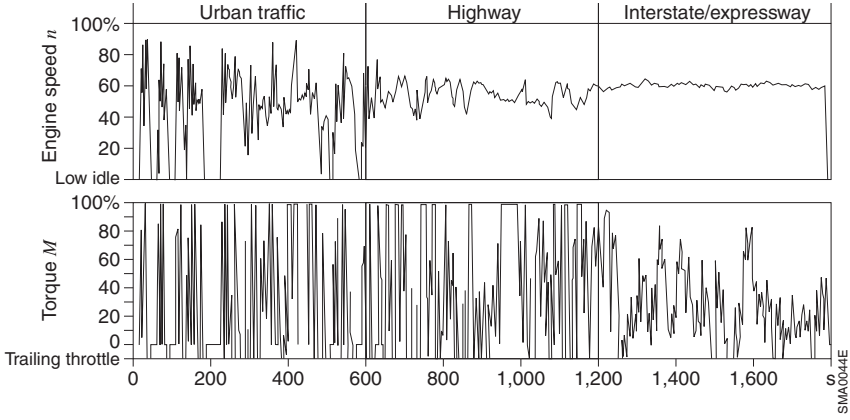


Figure 18: Transient driving cycle ETC (Europe)

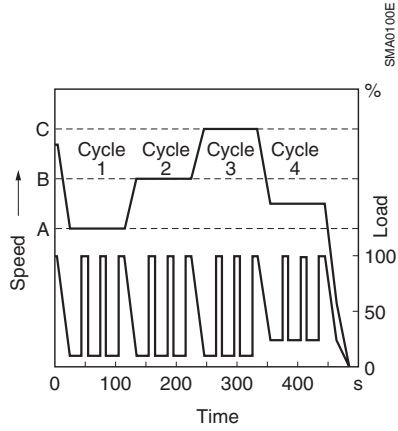


Europe
Up to Euro V all European test cycles started with a hot engine. Since EU VI a cold-start test and a hot-start test have been used

European Steady-State Cycle
For vehicles over 3.5 t permitted gross vehicle weight and more than 9 seats, the 13-stage test ESC (European Steady-State Cycle, Figure 17), has been in force in Europe since the introduction of Euro III (October 2000). The test procedure specifies measurements in 13 steady-state operating states calculated from the engine full-load curve. The emissions measured at each operating point are weighted according to certain factors. This also applies to power output. The test results are obtained for each pollutant by calculating the total of the weighted emissions divided by the total of the weighted power output.

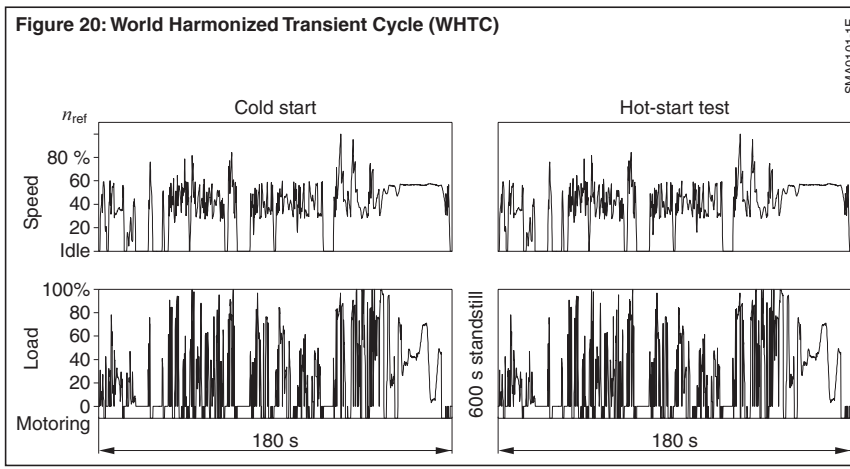
An additional three NO_x tests may be performed in the test range when certification is performed. The NO_x emissions may not vary by a significant degree from the levels measured at the adjacent operating points. The additional measuring has the goal of preventing engine modifications performed specially for the test.

Figure 19: European Load Response Test (ELR)



European Transient Cycle
As well as Euro III, the ETC (European Transient Cycle, Figure 18) was also introduced to determine gaseous emissions and particulate, and the ELR (European Load Response, Figure 19) test to measure exhaust-gas opacity. Under the Euro III standards, the ETC applies only to commercial vehicles with exhaust-gas treatment (particulate filter, NO_x reduction measures with NO_x storage catalytic converter or active SCR system); starting with Euro IV (October 2005) it has been obligatory for all vehicles.

Figure 20: World Harmonized Transient Cycle (WHTC)



The test cycle is derived from realistic road-driving patterns and is subdivided into three sections: an urban section, an extra-urban section, and an expressway section. The length of the test is 30 minutes, and the periods of time for which engine speeds and torque levels must be maintained are specified in seconds.

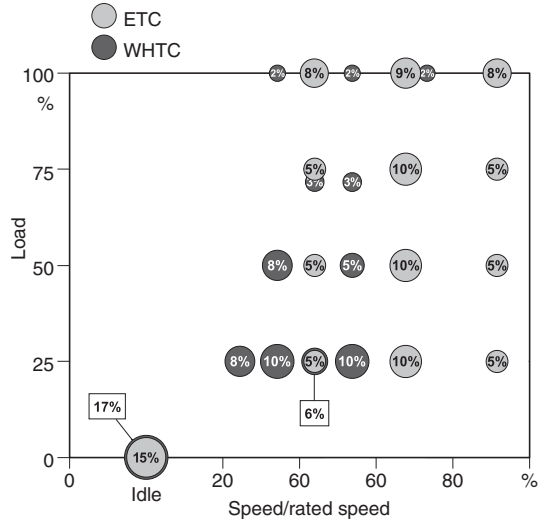
World harmonized cycles
Since 2013 world harmonized engine test cycles have had to be applied with the introduction of the Euro VI emission-limit level. The prescribed limits must be equally met both in the WHSC (World Harmonized Stationary Cycle) and in the WHTC (World Harmonized Transient Cycle, Figure 20). A new feature is a WNTE zone (World Harmonized Not To Exceed Zone), as was previously customary only in the USA. The NTE test is conducted in any driving mode within a specified engine-speed/torque range.

In relation to the WHTC limits WNTE emissions may exhibit an increase for NO_x (by 30 %) and particulates (by 60 %).

Compared with the current European tests, the harmonized engine tests are conceived in the direction of lower loads with fewer full-load operating points (Figure 21) and during the transient test clearly more overrun phases. The associated lower exhaust-gas temperatures pose a challenge to active exhaust-gas treatment systems, which must be regularly regenerated.

Real Road PEMS Tests
With the Euro VI legislation, in addition to the certification tests on the engine test bench, emissions from real road journeys are determined both for type approval and for checking vehicles already in operation (ISC, In-Service Conformity). Commercial vehicles are equipped with a mobile measuring device (PEMS, Portable Emissions Measurement System) for this purpose. The procedure is comparable with the RDE test of passenger cars and light commercial vehicles.

Figure 21: Comparison of ETC and WHTC test cycles



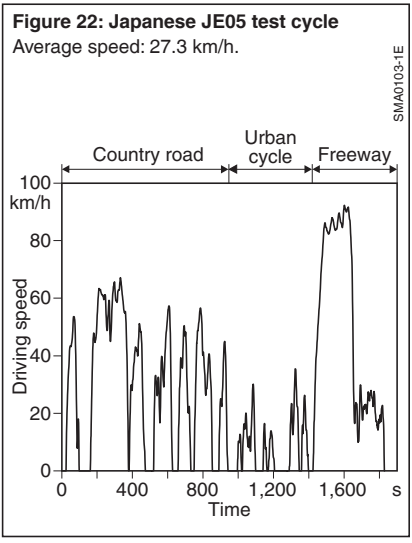
Japan

JE05 test cycle

Pollutant emissions were determined between October 2005 and October 2016 in the transient JE05 test cycle (Figure 22). Similarly to the European transient test for passenger cars, the JE05 test cycle for commercial vehicles comprises an extra-urban part, an urban part, and an expressway part. The test lasts 1,830 seconds and is started when the engine is hot.

Unlike the European and US commercial-vehicle tests, the JE05 test specifies the driving speed instead of the engine speed and engine torque. Because the test is conducted on an engine test bench, the quantities required of engine speed and torque are determined from the specified speeds and from the individual vehicle data using a conversion program. Required quantities include, among others, vehicle weight, tire rolling resistance, transmission ratios, torque curve, and maximum engine speed. With the introduction of the Post PNLT limit stages the harmonized European test cycles WHSC, WHTC, and WNTE are used for certification.

Figure 22: Japanese JE05 test cycle
Average speed: 27.3 km/h.



Emission legislation for motorcycles

The USA also led the way in introducing emission regulations for motorcycles. But because motorcycles only constitute a very small percentage of the overall traffic volume and consequently a very small percentage of the overall emissions and immissions, these requirements were for a long time not subject to further development compared with those for passenger cars and commercial vehicles. It was therefore left to the EU and Asian countries such as Japan to drive forward requirements for motorcycles and other L-category vehicles (i.e., two-, three-, and four-wheel vehicles such as mopeds, three-wheel delivery vans, minicars, and quad bikes). The EU regulations were incorporated into ECE Regulations, which are used by countries where two- and three-wheel vehicles are popular such as India, China, and other Southeast Asian countries as the basis for their national legislation. In these countries motorcycles play an important role as a means of transport for persons and goods, with high mileages in both urban and rural areas. Japan, Taiwan, and Thailand have introduced their own regulations for motorcycles.

The targets and instruments such as emission limits, test cycles, test procedures, and monitoring options are basically the same as those for passenger cars and light commercial vehicles (see Emission-control legislation, Overview).

Only the requirements for motorcycles (category L3/L4) with gasoline engines are considered here.

US legislation

Exhaust-gas emissions

The EPA introduced the first requirements for motorcycles in 1978. California had between 1988 and 2003 its own requirements, which were extended and tightened from 2004/2008. The EPA introduced only in 2006/2010 new requirements which are based on the Californian requirements from 2004/2008 and carry over to the entire USA.

Motorcycles are subdivided into on-highway and off-highway, each with

their own legislation, and encompass two- and three-wheel vehicles. Only on-highway motorcycles are considered here. These have since 2004 (CARB) and since 2006 (EPA) been classified according to their piston displacement into four classes to which the different durability requirements regarding mileage apply:
– Class I-A: < 50 ccm, 5 years/6,000 km,
– Class I-B: 50 to 169 ccm, 5 years/12,000 km,
– Class II: 170 to 279 ccm, 5 years/18,000 km,
– Class III: ≥ 280 ccm, 5 years/30,000 km.

The exhaust-gas emissions are measured in the FTP cycle on a chassis dynamometer at a temperature of between 20 °C and 30 °C. The emission limits are correlated with the route ridden during the test and are expressed in grams per km. The limits depend on the class and are not fuel-dependent.

Class I-A / I-B and Class II from 2004 (CARB) and from 2006 (EPA):
– carbon monoxide (CO) 12.0 g/km,
– hydrocarbons (HC) 1.0 g/km.

Class III 2004 – 2007 (CARB) and 2006 – 2009 (EPA Tier 1):
– carbon monoxide (CO) 12.0 g/km,
– hydrocarbons (HC) + nitrogen oxides (NO_x) 1.4 g/km.

Class III from 2008 (CARB) and from 2010 (EPA Tier 2):
– carbon monoxide (CO) 12.0 g/km,
– hydrocarbons (HC) + nitrogen oxides (NO_x) 0.8 g/km.

For Class III the manufacturer can average the aggregate limit value for HC + NO_x over its sold fleet.

A manufacturer can optionally certify Class I and Class II vehicles in accordance with the requirements for Class III as well.

There are no requirements with regard to particulate emissions. Unlike passenger cars, there are also no further exhaust-gas requirements.

Evaporative emissions
CARB introduced requirements for evaporative emissions from 1983. Since 2001 all classes have been governed by a limit value of 2.0 g per test in the SHED test (cf. Exhaust-gas measuring techniques), which differs in terms of procedure and duration from the test for passenger cars. Here, after the vehicle has undergone preconditioning including preparation of the activated carbon canister and with the tank filled to the defined level of 40 %, the actual tests start.

1st test: Tank-breathing losses
A linear temperature increase is conducted over 60 minutes in the SHED. Here a distinction is made between exposed and non-exposed tank and between fuel-liquid and fuel-vapor temperature. The hydrocarbon emissions given off by the vehicle are measured.

2nd test: Hot-soak losses
Following the first test the vehicle is heated in a defined riding cycle and then the increase in HC concentration is measured in the SHED over a period of one hour as the vehicle cools down.

The limit value applies to the sum total of the HC emissions in the two tests.

The EPA does not demand a SHED test, but instead focuses on limiting the permeability for hydrocarbons from fuel tank and lines by means of permeation limits. To this end, the vehicle manufacturer can either carry out a permeation test, use EPA-certified materials (defined permeation factors), or install certified components from suppliers.

Permeation limits:
– Fuel tank ≤ 1.5 g/m² and day,
– Fuel line ≤ 15 g/m² and day.

There are not requirements for consumption measurement or greenhouse-gas emissions for motorcycles.

CARB is working on an update of the legislation including OBD for motorcycles and checking a harmonization with the EU requirements. EPA is following the development at CARB and within the

framework of UN ECE and is planning from Spring 2021 a reassessment of the motorcycle regulations.

EU/ECE legislation

The basis of the EU emission legislation for motorcycles and other L-category vehicles with two, three and four wheels is the General Type Approval Directive 92/61/EEC [10]. With Directive 97/24/EC [11] limits and exhaust-gas emissions were defined on a European level for the first time and have been constantly updated ever since.

As for passenger cars and light commercial vehicles, so too for L-category vehicles the EU legislation serves as the model for the UN Regulations. These feature all the contents of the EU legislation, i.e., they reflect the EU stages described below. New requirements, which are aiming for worldwide harmonization and acceptance, are elaborated on a UN level. This applies in particular to the Worldwide harmonized Motorcycle Testing Cycle (WMTC) in the Global Technical Regulation GTR No. 2 [12].

The legislative stages for motorcycles are designated Euro or EU and denote an ever-further tightening of the exhaust-gas limits in the Type I test:
- Euro 1 (from 1999),
- Euro 2 (from 2003),
- Euro 3 (from 2006),
- Euro 4 (from 2016),
- Euro 5 (from 2020).

As for passenger cars, a type approval applies to an EU stage up to the point when the next stage is prescribed with binding effect.

The exhaust-gas emissions are measured in a defined test cycle on a chassis dynamometer in the Type I test at a temperature of between 20 °C and 30 °C. The emission limits are correlated with the route ridden during the test and are expressed in milligrams per km. They are in part fuel-dependent, i.e., different for gasoline and diesel engines.

For motorcycles, the test cycle or its applicable parts and their weighting are determined, depending on parameters like piston displacement and top speed. Here, from Euro 3 the synthetic cycle ECE-R40

Table 4: Emission limits Euro 3 to Euro 5 for Class L3 vehicles with gasoline engines (PM only for direct-injection gasoline engine)

Stage	Classification	Cycle/ weighting	CO mg/km	HC mg/km	NMHC mg/km	NO_x mg/km	PM mg/km
Euro 2	Displacement < 150 cm^3	R40 UDC	5,500	1,200	–	300	–
	Displacement ≥ 150 cm^3	R40 UDC	5,500	1000	–	300	–
Euro 3	Displacement < 150 cm^3	R40 UDC	2,000	800	–	150	–
	Displacement ≥ 150 cm^3	R40 UDC + EUDC	2,000	300	–	150	–
	v_{max} < 130 km/h	WMTC	2,620	750	–	170	–
	v_{max} ≥ 130 km/h	WMTC	2,620	330	–	220	–
Euro 4	v_{max} < 130 km/h	WMTC 30:70	1,140	380	–	70	–
	v_{max} ≥ 130 km/h	WMTC 25:50:25	1,140	170	–	90	–
Euro 5	v_{max} < 130 km/h	WMTC 30:70	1000	100	68	60	4.5
	v_{max} ≥ 130 km/h	WMTC 25:50:25	1000	100	68	60	4.5

is superseded by the more realistic dynamic cycle WMTC.

The EU emission legislation specifies for vehicles with gasoline engines limits for the following pollutants:
– Carbon monoxide (CO),
– Hydrocarbons (HC), from Euro 5 also for non-methane hydrocarbons (NMHC),
– Nitrogen oxides (NO_x),
– Particulate matter (PM) (from Euro 5 for gasoline direct-injection engines).

The emission limits for motorcycles (L3) and the test cycles are shown in Table 4. Nominally the Euro 5 limits match the limits for passenger cars Euro 5b.

To obtain approval for each vehicle model, the manufacturer must prove that the emissions of the limited pollutants do not exceed the respective limits over the service life prescribed by law (as distance traveled in kilometers). From Euro 4 for motorcycles with a top speed of 130 km/h or more the service life is 35,000 km, for a top speed of less than 130 km/h it is 20,000 km.

The type-approval test in the Type I test is conducted with a test vehicle which has been broken in for 1,000 km (from Euro 5 2,500 km or 3,500 km for top speed ≤ or > 130 km/h). Since the emission limits apply to the vehicle at the end of its service life, the measured values are multiplied by deterioration factors (from Euro 5 also additive factors) and then compared with the limit values. For this purpose either legally prescribed deterioration factors are used or the manufacturer determines these specially for this vehicle type in the Type V test (vehicle endurance run or from Euro 5 component aging).

Aside from the primary requirement in the Type I test, there is as a further exhaust-gas requirement the Type II test, in which the CO emissions are determined at idle.

As for passenger cars, there is no Type VI test at –7 °C for motorcycles.

OBD I is introduced with Euro 4, OBD II in two stages as of Euro 5.

Evaporative emissions
As with passenger cars, motorcycles with gasoline engines are subject to requirements to limit evaporative emissions (see Exhaust-gas measuring techniques) in the Type IV test. Evaporative emissions are determined from Euro 4 in the SHED (for L3 motorcycles). The procedure and the duration differ from the test for passenger cars and correspond to the Californian SHED test (see Evaporative emissions CARB).

The limit value applies to the sum total of the HC emissions in the two tests (tank-breathing and hot-soak losses):
– 2 g/test for Euro 4 (from 1/2016),
– 1.5 g/test for Euro 5 (from 1/2020).

CO_2 emissions
For motorcycles, the CO_2 emissions must also be measured in the same test as for the Type I test and from them the fuel consumption calculated. There are however no fleet targets as for passenger cars and light commercial vehicles.

Japanese legislation
The permissible emissions are being reduced in stages in Japan (similarly to in the EU).

The Japan-specific legislation with a synthetic cycle was replaced from 2012 on the basis of GTR No. 2 by the WMTC, with limit values equivalent to the previous stage.

The next WMTC stage applies from 10/2016 to new domestic types and from 9/2017 to all types and to imports.

A new requirement added from 10/2017 is the limitation of evaporative emissions. The test procedure corresponds to the Californian SHED test with a limit value of 2.0 g HC per test.

From 10/2017 Japan will also introduce its own J-OBD I regulation. This is based on Euro 4-OBD I and on the general OBD requirements of CARB-OBD II. J-OBD II on the basis of Euro 5 is planned.

For motorcycles there are no requirements with regard to consumption measurement or to CO_2 emissions.

Chinese legislation
China has introduced the EU legislation in stages. The stage China 3 from 7/2008 will be replaced from 7/2018 by China 4, which corresponds in terms of exhaust-gas and evaporative emissions to Euro 4. Euro 4-OBD I on the other hand has not been adopted; instead, a China-specific basic OBD I regulation is being developed. The next stage China 5 based on Euro 5 is in development and is to be published in 2022.

Indian legislation
India has introduced the EU legislation in stages. The stage Bharat IV from 4/2016 corresponds to Euro 3. The stage Euro 4 is to be skipped and from 4/2020 replaced by Bharat VI, which corresponds in terms of exhaust-gas and evaporative emissions to Euro 5. OBD I is to be introduced in 4/2020 and OBD II from 4/2023, if necessary in two stages as in the EU.

References
[1] 70/220/EEC: Council Directive (of the European Communities) of 20 March 1970 on the approximation of laws of the Member States relating to measures to be taken against air pollution by gases from positive-ignition engines of motor vehicles.
[2] Regulation (EC) No. 715/2007 of the European Parliament and of the Council of 20 June 2007 on type approval of motor vehicles with respect to emissions from light passenger and commercial vehicles (Euro 5 and Euro 6) and on access to vehicle repair and maintenance information.
[3] UN GTR No. 15: Global Technical Regulation No. 15 on Worldwide harmonized Light vehicles Test Procedure (ECE/TRANS/180/Add.15).
[4] UN GTR No. 19: Global Technical Regulation No. 19 – EVAPorative emission test procedure for the Worldwide harmonized Light vehicle Test Procedure (WLTP EVAP) (ECE/TRANS/180/Add.19).
[5] Regulation (EC) No. 595/2009 of the European Parliament and of the Council of 18 June 2009 on type approval of motor vehicles and engines with respect to emissions from heavy commercial vehicles (Euro VI) and on access to vehicle repair and maintenance information, and amending Regulation (EC) No. 715/2007 and Directive 2007/46/EC and repealing Directives 80/1269/EEC, 2005/55/EC and 2005/78/EC.
[6] 88/77/EEC: Council Directive of 3 December 1987 on the approximation of the laws of the Member States relating to measures to be taken against the emission of gaseous pollutants from diesel engines for use in vehicles.
[7] Commission Regulation (EU) No. 64/2012 of 23 January 2012 amending Regulation (EU) No. 582/2011 implementing and amending Regulation (EC) No. 95/2009 of the European Parliament and of the Council with respect to emissions from heavy commercial vehicles (Euro VI) (Text with EEA relevance).

[8] Commission Regulation (EU) 2016/1718 of 20 September 2016 amending Regulation (EU) No. 582/2011 with respect to emissions from heavy commercial vehicles as regards the provisions on testing with portable emission measurement systems (PEMS) and the procedure for the testing of the durability of replacement pollution control devices (Text with EEA relevance).
[9] Commission Regulation (EU) 2020/1181 of 7 August 2020 correcting certain language versions of Directive 2007/46/EC of the European Parliament and of the Council establishing a framework for the approval of motor vehicles and their trailers, and of systems, components and separate technical units intended for such vehicles (Framework Directive), correcting certain language versions of Commission Regulation (EU) No. 582/2011 implementing and amending Regulation (EC) No 595/2009 of the European Parliament and of the Council with respect to emissions from heavy-duty vehicles (Euro VI) and amending Annexes I and III to Directive 2007/46/EC of the European Parliament and of the Council, and correcting the Danish language version of Commission Regulation (EU) 2017/2400 implementing Regulation (EC) No. 595/2009 of the European Parliament and of the Council as regards the determination of the CO2 emissions and fuel consumption of heavy-duty vehicles and amending Directive 2007/46/EC of the European Parliament and of the Council and Commission Regulation (EU) No. 582/2011 (Text with EEA relevance).
[10] Council Directive 92/61/EEC of 30 June 1992 relating to the type-approval of two or three-wheel motor vehicles.
[11] Directive 97/24/EC of the European Parliament and of the Council of 17 June 1997 on certain components and characteristics of two or three-wheel motor vehicles.
[12] UN GTR No. 2: Measurement procedure for two-wheeled motorcycles equipped with a positive or compression ignition engine with regard to the emission of gaseous pollutants, CO2 emissions and fuel consumption (ECE/TRANS/180/Add.2).
[13] K. Reif (Editor): Ottomotor-Management – Bosch Fachinformation Automobil. 4th Edition, Springer Vieweg, 2014.

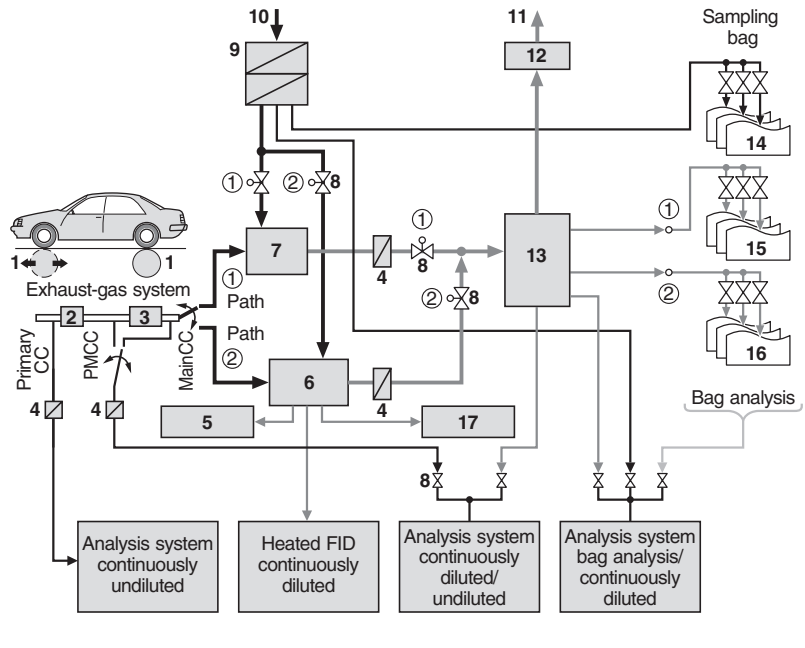
Exhaust-gas measuring techniques

Exhaust-gas testing

Requirements
Exhaust-gas tests on chassis dynamometers are used for the type approval to attain General Certification, as well as to develop engine or other components. They differ from exhaust-gas tests that are conducted in Germany for example in the course of “general and partial inspection for exhaust gas” (§29 StVZO, [1]) using workshop measuring devices. In addition, exhaust-gas tests are carried out on engine test benches, for instance for the type approval of heavy commercial vehicles.

Provided for information purposes only, no liability assumed for completeness!

Figure 1: Exhaust-gas test on the chassis dynamometer
1 Roller with dynamometer, 2 Upstream catalytic converter, 3 Underfloor catalytic converter, 4 Filter, 5 Particulate filter, 6 Dilution tunnel, 7 Mix-T, 8 Valve, 9 Dilution-air conditioner, 10 Dilution air, 11 Exhaust-gas/air mixture, 12 Blower, 13 CVS system (Constant Volume Sampling), 14 Dilution-air sample bag, 15 Exhaust-gas sample bag (from Mix-T), 16 Exhaust-gas sample bag (from dilution tunnel), 17 Particulate counter.
① Path for exhaust-gas measurement via Mix-T (without determination of particulate emission),
② Path for exhaust-gas measurement via dilution tunnel (with determination of particulate emission).



The exhaust-gas test on chassis dynamometers is carried out on vehicles. The methods used are defined to simulate actual vehicle operation on the road as far as possible. Measurement on a chassis dynamometer offers the following advantages here:
– Highly reproducible results, as environmental conditions can be kept constant.
– Good comparability of tests, as a defined speed-time profile can be driven independently of traffic flow.
– Stationary setup of the measuring techniques required.

Test setup

General setup

The test vehicle is parked on a chassis dynamometer with its drive wheels on the rollers (Figure 1). This means that the forces acting on the vehicle, i.e., the vehicle's moments of inertia, rolling resistance and aerodynamic drag, must be simulated so that the trip on the test bench reproduces emissions comparable to those obtained during an on-road trip. For this purpose, asynchronous machines, direct-current machines, or even electrodynamic retarders on older test benches, generate a suitable speed-dependent load that acts on the rollers for the vehicle to overcome. More modern machines use electric flywheel simulation to reproduce this inertia. Older test benches use real flywheels of different sizes attached by rapid couplings to the rollers to simulate the vehicle weight. A blower mounted in front of the vehicle provides the necessary engine cooling.

The test-vehicle exhaust pipe is generally a gas-tight attachment to the exhaust-gas collection system – the dilution system is described below. A proportion of the exhaust gas is collected there. At the end of the driving test, the gas is analyzed for gaseous emission-limit components (hydrocarbons, nitrogen oxides and carbon monoxide) and carbon dioxide (to determine fuel consumption).

Following the introduction of the emission-control legislation, particulate emissions were initially limited in diesel-engine vehicles only. In the last few years, legislators have also begun to limit these emissions in vehicles with gasoline engines. To determine particulate emissions, a "dilution tunnel" is used with a high internal flow turbulence (Reynolds number > 40,000). Particulate filters are also used to calculate particulate emission based on load.

In addition, and for development purposes, part of the exhaust gas flow can be extracted continuously from sampling points in the vehicle's exhaust-gas system or dilution system to analyze the pollutant concentrations.

The test cycle is repeated by a driver in the vehicle. The required and current driving speeds are displayed on a driver control-station monitor. In some cases, an automated driving system replaces the driver to increase the reproducibility of test results.

Test setup for diesel-engine vehicles

To determine the pollutant emissions from diesel vehicles, it is necessary to make some changes to the test-bench setup and to the measuring techniques used. The complete sample-taking system, including the exhaust-gas measuring device for hydrocarbons, must be heated to 190 °C. This is to prevent condensation of hydrocarbons which have high boiling points, and to evaporate the hydrocarbons that have already condensed in the diesel exhaust gas.

Dilution system

Purpose of the CVS method

The most commonly used method of collecting the exhaust gases emitted from an engine is the CVS dilution procedure (Constant Volume Sampling). It was introduced for the first time in the USA in 1972 for passenger cars and light-duty commercial vehicles. In the meantime, it has been improved in several stages. The CVS method is used in other countries, such as Japan. It has also been in used in Europe since 1982. It is therefore an exhaust-gas collection method that is recognized throughout the world.

In the CVS method, the exhaust gas is only analyzed at the end of the test. Therefore the condensation of water vapor and the resulting nitrogen-oxide losses and

secondary reactions in the collected exhaust gas must be avoided.

Principle of the CVS method
The exhaust gases emitted by the test vehicle are diluted in the Mix-T or in the dilution tunnel with ambient air at a mean ratio of 1:5 to 1:10, and extracted using a special system of pumps in such a way that the total volumetric flow composed of exhaust gas and dilution air is constant. The admixture of dilution air is therefore dependent on the momentary exhaust-gas volumetric flow. A representative sample is continuously extracted from the diluted exhaust-gas flow and is collected in one or more exhaust-gas sample bags. The sampling volumetric flow is constant during the bag-filling phase. Therefore, the pollutant concentration in a sample bag at the end of the filling process is identical to the mean value of the concentration in the diluted exhaust gas during the bag-filling process.

While the exhaust-gas sample bags are being filled, a sample of the dilution air is taken and collected in one or more air sample bags in order to measure the pollutant concentration in the dilution air.

Filling the sample bags generally corresponds to the phases in which the test cycles are divided (e.g., the ht phase in the FTP 75 test cycle).

The pollutant mass emitted during the test is calculated from the total volume of the diluted exhaust gas and the pollutant concentrations in the exhaust-gas and air-sample bags.

Dilution systems
There are two alternative methods for achieving the constant volumetric flow in the diluted exhaust gas:
– PDP method (Positive Displacement Pump): A rotary-piston blower (Roots blower) is used.
– CFV method (Critical Flow Venturi): A venturi tube and a standard blower are used in the critical state.

Advances in the CVS method
Diluting the exhaust gas causes a reduction in pollutant concentrations as a factor of the dilution. The concentrations of some pollutants (especially hydrocarbon compounds) in the diluted exhaust gas are comparable to the concentrations in the dilution air (or lower) in certain test phases, since pollutant emissions have been significantly reduced in recent years as emission limits have become more stringent. This poses a problem from the measuring-process aspect as the difference in the two values is crucial for the pollutant emissions. A further challenge is presented by the precision of the measuring devices used to analyze the pollutants.

The following measures have been generally implemented in order to counter the problems described above:
– Lowering the dilution; this requires precautions to prevent water from condensing, e.g., by heating sections of the dilution systems, or drying or warming dilution air on vehicles with a gasoline engine.
– Reducing and stabilizing pollutant concentrations in the dilution air, e.g., by using activated charcoal filters.
– Optimizing the measuring devices (including dilution systems), e.g., by appropriately selecting or preconditioning the materials used and system setups or by using modified electronic components.
– Optimizing processes, e.g., by applying special purge procedures.

Bag Mini Diluter
As an alternative to the improvements in CVS technology described above, a new type of dilution system was developed in the USA: the Bag Mini Diluter (BMD). Here, part of the exhaust gas flow is diluted at a constant ratio with dried, heated, pollutant-free zero gas (e.g., cleaned air). During the test, part of this diluted exhaust-gas flow that is proportional to the exhaust-gas volumetric flow is filled in exhaust-gas sample bags and analyzed at the end of the driving test.

In this procedure, dilution is performed with a pollutant-free zero gas free of pollutants and not with air containing pollutants. This has the purpose of avoiding the air-sample bag analysis and the subsequent differential formation of exhaust-gas and air-sample bag concentrations. However, a more complex procedure is required than that for the CVS method, e.g., one requirement is to determine the (undiluted) exhaust-gas volumetric flow and the proportional sample-bag filling.

Table 1: Discontinuous measuring methods

Components	Procedure
CO, CO_2	Non-dispersive infrared analyzer (NDIR)
Nitrogen oxides (NO_x)	Chemiluminescence detector (CLD)
Total hydrocarbon (THC)	Flame ionization detector (FID)
CH_4	Combined design of gas chromatographic procedure and flame ionization detector (GC FID)
CH_3OH, CH_2O	Combined design of impinger or cartridge process and chromatographic analysis techniques; mandatory in the USA when certain fuels are used
Particulates	1.) Gravimetric method: Weighing of particulate filters before and after the test drive 2.) Particulate counting

Exhaust-gas measuring devices

In vehicles with gasoline engines, the emissions of limited gaseous pollutants are calculated from the concentrations in exhaust-gas and air sample bags (CVS method). Emission-control legislation defines globally standard test procedures for this purpose (Table 1).

Essentially, the same devices are used to measure the concentrations of gaseous pollutants in the exhaust gas of gasoline-engine vehicles as for diesel-engine vehicles. There is however a difference in how the hydrocarbon emissions (HC) are determined. This is done not in the exhaust-gas sample bag, but through continuous analysis of a partial flow of diluted exhaust gas and integration of the measured concentration via the driving test. The reason for this is that the hydrocarbons (which have a high boiling point) condense in the (non-heated) exhaust-gas sample bag.

For development purposes, many test benches also include the continuous measurement of pollutant concentrations in the vehicle exhaust-gas system or the dilution system. The reason is to capture data for the components under control, as well as for other components not subject to legislation. Other test procedures than those listed in Table 1 are required for this, e.g.:

- Paramagnetic method (to measure O_2 concentration).
- Cutter FID: a combination of flame ionization detector and absorber for non-methane hydrocarbons (to measure the CH_4 concentration).
- Mass spectroscopy (multi-component analyzer).
- FTIR spectroscopy (Fourier Transform Infrared, multi-component analyzer).
- IR laser spectroscopy (multi-component analyzer).

A description of the main measuring devices is given below.

Test-bench measuring techniques

NDIR analyzer

The NDIR (Non-Dispersive Infrared) analyzer utilizes the property of certain gases to absorb infrared radiation within a narrow wavelength range. Absorbed radiation is converted to vibration or rotation energy by the absorbing molecules. In turn, this energy can be measured as heat. The phenomenon described occurs in molecules that are formed from atoms of at least two different elements, e.g., CO, CO_2, C_6H_{14} or SO_2.

There are a number of variants of NDIR analyzers; the main component parts are a source of infrared light (Figure 2), an absorption cell (cuvette) through which the test gas is routed, a reference cell generally positioned in parallel (filled with inert gas, e.g., N_2), a rotating chopper, and a detector. The detector comprises two chambers connected by a membrane and containing samples of the gas components under analysis. Radiation from the reference cell is absorbed in one chamber and radiation from the cuvette in the other. Radiation from the cuvette may have already been reduced by absorption in the test gas. The difference in radiant energy causes a flow movement that is measured by a flow sensor or a pressure sensor. The rotating chopper interrupts the infrared radiation in cycles, causing the flow movement to change direction and therefore a modulation of the sensor signal.

NDIR analyzers possess strong cross sensitivity to water vapor in the test gas since H_2O molecules absorb a wide range of infrared radiation wavelengths. This is the reason why NDIR analyzers are positioned downstream of a test-gas treatment system (e.g., a gas cooler) to dry the exhaust gas when they are used to make measurements on undiluted exhaust gas.

Chemiluminescence detector (CLD)

In a reaction chamber, the test gas is mixed with ozone that is produced from oxygen in a high-voltage discharge (Figure 3). The nitrogen monoxide content in the test gas oxidizes to nitrogen dioxide in this environment; some of the molecules produced are in a state of excitation. When these molecules return to their basic state, energy is released in the form of light (chemiluminescence). A detector (e.g., photomultiplier) measures the emitted luminous energy; under specific conditions, it is proportional to the nitrogen-monoxide concentration (NO) in the test gas.

Figure 2: Test chamber acc. to the NDIR method

1 Gas outlet, 2 Absorption cell,
3 Test-gas inlet, 4 Optical filter,
5 Infrared light source, 6 Infrared radiation,
7 Reference cell, 8 Rotating chopper,
9 Detector.

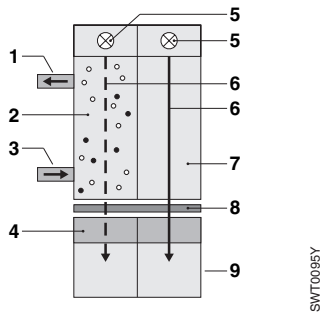
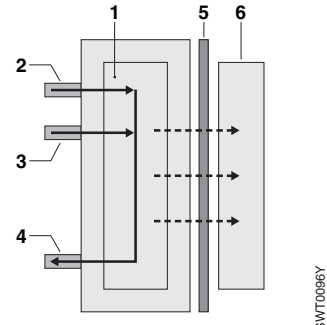


Figure 3: Design of chemiluminescence detector

1 Reaction chamber, 2 Ozone inlet,
3 Test-gas inlet, 4 Gas outlet, 5 Filter,
6 Detector.

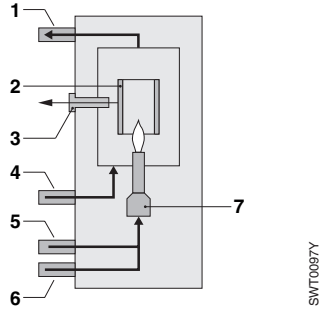


It is a requirement to measure the NO and NO_2 molecules as the legislation regularizes the emission of the total nitrogen oxides. However, since the test principle of the chemiluminescence detector is limited to measuring the NO concentration, the test gas is channeled through a converter that reduces the nitrogen dioxide to nitrogen monoxide.

Flame Ionization Detector (FID)
The test gas is burned in a hydrogen flame (Figure 4), where carbon radicals are formed and some of these radicals are temporarily ionized. The radicals are discharged at a collector electrode. The current produced is measured and is proportional to the number of carbon atoms in the test gas.

GC FID and Cutter FID
There are two generally common methods to measure the methane concentration (CH_4) in the test gas. Each method consists of the combination of a CH_4-separating element and a flame ionization detector. In these methods, either a gas-chromatography column (GC FID) or a heated catalytic converter oxidizing the non-CH_4 hydrocarbons (cutter FID) are used to separate the methane.

Figure 4: Design of flame ionization detector
1 Gas outlet, 2 Collector electrode, 3 Amplifier output, 4 Combustion air, 5 Test-gas inlet, 6 Combustion gas (H_2/He), 7 Burner.



Unlike the cutter FID, the GC FID can only determine the CH_4 concentration discontinuously (typical interval between two measurements: 30 to 45 s).

Paramagnetic detector (PMD)
There are different constructions of paramagnetic detectors (dependent on the manufacturer). The constructions are based on the phenomenon that forces with paramagnetic properties (such as oxygen) act on molecules in inhomogeneous magnetic fields. These forces cause the molecules to move. The movement is sensed by a special detector and is proportional to the concentration of molecules in the test gas.

Measuring particulate emission
In addition to gaseous pollutants, solid particulates are also measured, as they are also pollutants subject to legislation. Currently, the gravimetric process is the process specified by law to measure particulate emissions.

Gravimetric process (particulate-filter process)
Part of the diluted exhaust gas is sampled from the dilution tunnel (CVS method) during the driving test and then channeled through particulate filters. The particulate loading is calculated from the weight of the particulate filters before and after the test. The particulate emission during the driving test is then calculated from the load, the total volume of the diluted exhaust gas, and the partial volume channeled through the particulate filters.

The gravimetric process has the following disadvantages:
– Relatively high detection limit, only reducible to a limited extent by using complex instrumentation (e.g., to optimize the tunnel geometry).
– It is not possible to measure particulate emissions continuously.
– The process is complex as particulate filters have to be conditioned in order to minimize environmental influences.
– It is not possible to select the chemical composition of particulates or particulate size.

Particulate counting
Because of the disadvantages mentioned and the ongoing reduction of the limit values, increasingly the number of emitted particulates will be limited, as well as the particulate emission (particulate mass per distance traveled).

The "Condensation Particulate Counter" (CPC) has been earmarked as the measuring device for calculating the number of particulates in compliance with legislation (particulate counting). In this counter, a small partial flow of the diluted exhaust gas (aerosol) is mixed with saturated butanol vapor. The process of the butanol condensing on the solid particulates causes the particulates to increase dramatically in size so that it is possible to calculate the number of particulates in the aerosol with the aid of scattered-light measurement.

The number of particulates in the diluted exhaust gas is continuously calculated; integration of the measured values produces the number of particulates over the driving test.

Determination of particulate size distribution
There is increasing interest in acquiring knowledge of the size distribution of particulates in the exhaust gas of a vehicle. Examples of devices that supply this information are the Scanning Mobility Particle Sizer (SMPS), Electrical Low Pressure Impactor (ELPI), and Differential Mobility Spectrometer (DMS).

Portable emission measuring devices
A PEMS measuring device (Portable Emission Measurement Systems) consists primarily of three components to ensure a complete analysis of the vehicle emissions.

Gas PEMS
Gas-analysis apparatus is used to measure the pollutant emissions in the exhaust-gas flow. This consists of an NDUV analyzer (Non-Dispersive Ultraviolet analyzer) to determine NO and NO_2, an NDIR analyzer (see Test-bench measuring techniques) to measure CO and CO_2, and an O_2 sensor.

The NDUV analyzer is a multi-component UV photometer. The measuring principle is based here essentially on the characteristic attribute of NO and NO_2 to absorb luminous radiation in a wavelength range of 200 to 500 nm.

PN PEMS (Particle Number)
Two different measuring principles are used to measure the particulates. Firstly, portable "Condensation Particle Counters" (CPCs) are used which through condensation of the particulates, for example with n-butanol vapor make it possible to detect nanoscale particulates optically (size range: 23 to 700 nm) (see Test-bench measuring techniques, particulate counting).

Secondly, portable "Diffusion Chargers" (DCs) are used to count particulates. The measuring principle is based on the electrostatic precipitation of charged particulates. The particulates are charged by collisions with ions which are created by a gas discharge in a charging chamber in the measuring device. Precipitation of the charged particulates creates a potential difference between the metallic precipitation surface and the surrounding area. It is then possible by means of numerical methods to use this potential difference to determine theoretically the particulate number.

EFM (Exhaust Flow Meter)
The exhaust-gas volumetric flow is measured with an exhaust flow meter (EFM). Here the pressure loss is measured at a defined flow resistance inside the test pipe. The differential pressure can then be used to calculate the volume of exhaust gas.

The EFM is a test pipe mounted on the exhaust tailpipe which contains both pressure and temperature sensors. The sampling point for the gas-analysis apparatus and the sampling point for measuring the particulate number are also integrated on this adapter.

Testing commercial vehicles
In the USA the transient test method has been prescribed since model year 1987 for testing the emissions from diesel engines in heavy commercial vehicles over 8,500 lbs (EPA) and 14,000 lbs (CARB) gross vehicle weight. It is performed on dynamic engine test benches. In Europe this test method was introduced with Euro III for engines with active exhaust-gas-treatment (particulate filter or selective catalytic reduction, SCR) and with Euro IV for all engines in heavy commercial vehicles over 3.5 t gross vehicle weight. Since Euro V the test on the engine test bench is dependent no longer on the gross vehicle weight but on the reference weight (tare weight plus 100 kg).

The transient test method also uses the CVS method. However, the size of the engines demands a test setup with a substantially higher throughput in order to keep to the same dilution ratios as for cars and light-duty commercial vehicles. Double dilution (through a secondary tunnel) approved by legislators helps to limit the increased complexity of instrumentation. The diluted exhaust-gas volumetric flow can be set alternatively using a Roots blower or critical-flow venturi nozzles.

Another possibility would be to calculate the particulate emissions with a partial-flow dilution system, provided that the remaining pollutants in the undiluted exhaust gas are measured.

With Euro VI (2013) limit values for the particulate number are binding on all new model types. The value is $8 \cdot 10^{11}$ particles in steady-state and $6 \cdot 10^{11}$ particles per kWh in transient operation.

Diesel smoke-emission test

Methods
Separate legislation for testing the smoke emissions of diesel-engine vehicles came into force long before the introduction of legislation for testing gaseous pollutants. All existing smoke tests are coupled closely with the measuring devices used. One measure of smoke emission (soot, particulates) is the smoke number.

Two methods are essentially customary for measuring this value. In the absorption method (opacity measurement), the opacity of the exhaust gas is indicated by the degree to which it blocks the passage of a beam of light shining through it (Figure 5). In the filter method (measurement of reflected light), a specified quantity of exhaust gas is routed through a filter element. The degree of filter discoloration provides an indication of the amount of soot contained in the exhaust gas (smoke tester, Figure 7).

Measurement of diesel smoke emissions is relevant only if the engine is under load, since it is only when the engine is operated under load that emission of significant levels of particulates occur. Two different test procedures are also commonly used here. Firstly, measurements under full load, e.g., on a chassis dynamometer, or on a defined test circuit. Secondly, measurements under unrestricted acceleration using a defined throttle burst under the load of the engine flywheel (Figure 6).

As the results of testing for diesel smoke emissions vary according to both test procedure and type of load, they are not generally suitable for direct mutual comparisons.

Opacimeter (absorption method)
Exhaust-gas opacity denotes the attenuation of light by absorption, diffraction, scatter and reflection on the particulates contained in the exhaust gas.

A transmitter and a photodetector are mounted on the exhaust pipe for measurement of the full flow with the opacimeter. In partial-gas devices, the exhaust gas is routed through an exhaust-gas sample probe and heated lines by a pump into the

measuring chamber (Figure 5). Using a longer measuring chamber improves the device's detection limit.

During the unrestricted acceleration part of the exhaust gas is delivered to the measuring chamber. A light beam passes through the measuring chamber now filled with exhaust gas. The attenuation of lights is determined photoelectrically and indicated as opacity T in % or as an absorption coefficient k in m^{-1} (Figure 6). A precisely defined measuring-chamber length and keeping the optical window clear of soot (by air curtains, i.e., tangential air flows) are basic requirements for high levels of accuracy and reproducibility of the measurements.

During testing under load, measurement and display are a continuous process. The opacimeter automatically determines the maximum value and calculates the mean from several gas pulses.

Smoke tester (filter method)

During soot measurement according to Bosch a defined exhaust-gas volume (330 cm^3) is drawn by means of a hand pump through a white filter paper. The blackening of the paper is determined by means of a reflex photometer. The degree of paper blackening is divided into numbers ranging between 0 and 10 (smoke number according to Bosch,

Figure 6: Gas-pulse measurement with the opacimeter

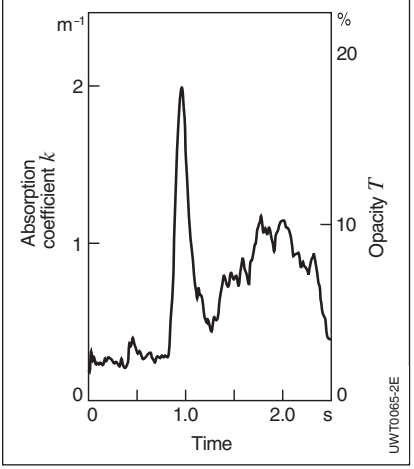
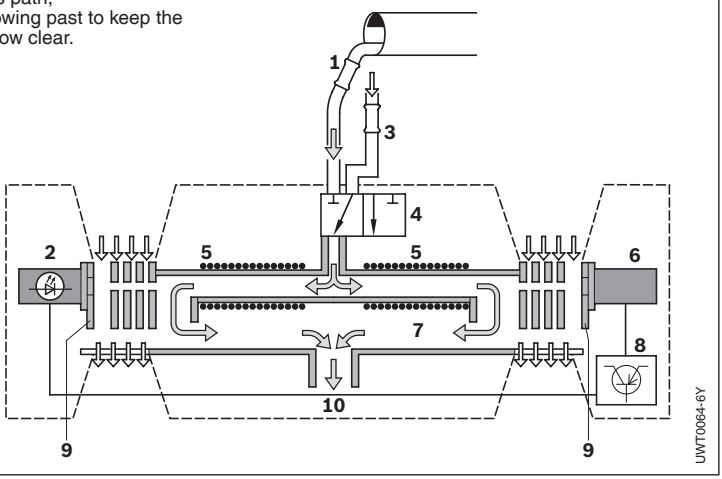


Figure 5: Opacimeter (absorption method)

1 Sample probe, 2 Light source, 3 Purge air for calibration, 4 Calibrating valve, 5 Heater, 6 Receiver, 7 Measuring chamber, 8 Electronic analyzer and display, 9 Optical window, 10 Measuring-chamber outlet, to pump.

➧ Exhaust-gas path,

▷ Purge air flowing past to keep the optical window clear.



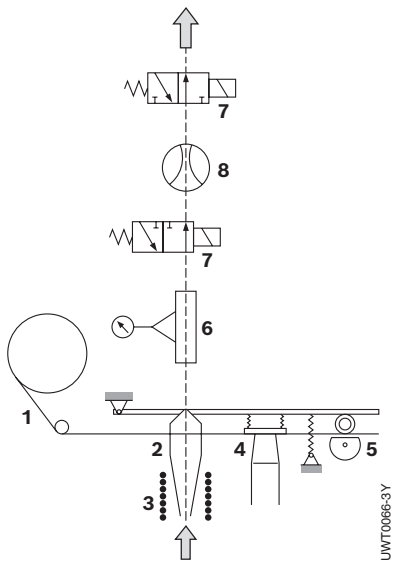
Bosch number), where 0 denotes white, unused paper and 10 denotes fully blackened measuring points.

An empirical correlation can be used to convert to soot-mass concentration in mg/m^3.

The method has been developed over decades. Thus, the hand pump, for example, was replaced by a continuously operating pump, the dead volume between the sample probe and the filter paper is taken into consideration, and the devices are heated to prevent condensation (Figure 7).

The filter smoke number (FSN) is defined in ISO 10054 [2]. It corresponds to the Bosch number for an effective suction length of 405 mm at a temperature of 298 K and a pressure of 1 bar.

Figure 7: Smoke tester (filter method)
1 Filter paper, 2 Gas passage,
3 Heater, 4 Reflex photometer,
5 Paper transport, 6 Volume measurement,
7 Purge-air changeover valves,
8 Pump.



Evaporative-emission test

Independently of the combustion pollutants produced in the engine, a gasoline-engine vehicle emits additional quantities of hydrocarbons (HC) through evaporation of fuel from the fuel tank and fuel system. The quantity evaporated is dependent on the design of the vehicle and on the fuel temperature. An activated carbon canister, in which the fuel vapors are stored, is generally used as a limiting measure. Since it only has a limited absorption capacity, the activated carbon canister must be regenerated at certain intervals. This is done by purging with fresh air while driving. The air/fuel mixture is delivered to the induction tract and combusted in the engine (see Evaporative-emissions control system). Many countries (e.g., the USA and Europe) have regulations which limit these evaporative losses.

Test methods

These evaporative emissions are usually quantified with the aid of a hermetically sealed climate chamber, the SHED (Sealed Housing for Evaporative emissions Determination). For the test, HC concentrations are measured at the beginning and the end of the test with a flame ionization detector (FID), with the difference representing the evaporative losses.

The evaporative losses must – depending on the country – be measured in some or in all of the operating states below and satisfy emission limits:

– Evaporation that emerges from the fuel system when the vehicle is parked with the engine warm following operation: hot-soak test (EU, USA, and others).
– Evaporation emerging from the fuel system as a result of temperature changes in the course of the day: tank-breathing test or diurnal test (EU, USA, and others).
– Evaporative emissions while driving, e.g., due to permeation: running-loss test (USA only).

Test procedures (examples for passenger cars and light commercial vehicles)
Evaporations are measured in several phases during a detailed prescribed test procedure. After the vehicle has undergone preconditioning including preparation of the activated charcoal canister ad with the tank filled to the defined level of 40 %, testing starts.

First test: Hot-soak losses (hot soak)
To measure the evaporative emissions in this test phase, the vehicle is heated up before the test by running through the test cycle valid for the country concerned. The engine is then turned off when the vehicle is in the climate chamber. The increase in HC concentration is measured for a period of one hour as the vehicle cools.

The vehicle's windows and trunk lid must remain open during the test. This allows evaporative losses from the vehicle interior to be included in measurements.

Second test: Tank-breathing losses
For this test, a typical temperature profile for a warm summer day (maximum temperature for EU: 35 °C; for EPA: 35.5 °C; for CARB: 40.6 °C) is simulated inside the hermetically sealed climate chamber. The hydrocarbons emitted by the vehicle under these conditions are then collected.

In the USA, both a 2-day diurnal test (48 hours; in-use test) and a 3-day diurnal test (72 hours; certification) must be carried out (the highest value over the course of the day is used in each case), where the preconditioning runs in each case slightly differently. The aim of the 3-day diurnal test is to check the sufficient capacity of the activated carbon canister. The 2-day diurnal test serves to check whether the activated carbon canister was sufficiently purged during the preconditioning preceding this test.

EU legislation from Euro 3 provided for a 24-hour test. For Euro 6d the test procedure was revised and a new test cycle WLTC, a 48-hour diurnal test and further tests for checking durability were added. Before the test procedure the activated carbon canister is to be aged according to a defined process. For further information, see Emission-control legislation.

Third test: Running losses
The running-loss test is conducted prior to the hot-soak test. It is used to assess the hydrocarbon emissions generated during vehicle operation in the prescribed test cycles (1 x FTP 72, 2 x NYCC, 1 x FTP 72; see US test cycles).

Limits
EU legislation
The total of the measurement results from the first and second days of the 48-hour diurnal test, the hot-soak tests and a double so-called permeation factor measured in an upstream procedure produces the evaporative losses. The tank is checked in a separate test procedure with regard to permeation emissions. The factor calculated from this has a bearing of the final calculation of the evaporative emissions. This total must lie below the demanded limit value of 2 g of evaporated hydrocarbons for all the measurements.

USA
In the USA (legislation according to CARB LEV II and EPA Tier 2), the evaporative losses monitored in the running-loss test must remain below 0.05 g per mile. The limit values for the hot-soak losses and the tank-breathing losses are defined for Tier 2 and LEV II as follows:
– 2-day diurnal + hot soak:
  1.2 g (EPA) / 0.65 g (CARB),
– 3-day diurnal + hot soak:
  0.95 g (EPA) / 0.50 g (CARB).

These limits must be adhered to over 120,000 miles (EPA) or 150,000 miles (CARB). They have been introduced in stages since model year 2004 and have applied in full since model year 2007 (EPA) and 2006 (CARB). Since model year 2009, EPA has alternatively permitted certification in accordance with CARB limits and CARB regulations (harmonization).

For PC (Passenger Car) and LDT1 (Light-Duty Truck), which are certified in accordance with the CARB ZEV legislation, a lower limit value in the SHED test

of 0.350 g HC per test and additionally the "zero evaporative emissions" requirement apply. Practically no emissions of fuel (limit value: 0.054 g HC/test) are meant. To this end, the above-described "3-day diurnal + hot soak" is carried out in the "rig test", i.e., a setup of tank, fuel lines, activated carbon canister, and engine by arrangement between CARB and the vehicle manufacturer.

With LEV III (phase-in 2015 through 2022) and Tier 3 (phase-in 2017 through 2022) the requirements for limiting evaporative emissions of the ZEV legislation is extended to all vehicles. The manufacturers can either adhere to the above-described ZEV limit values or alternatively the combination of slightly stricter limit values in the SHED test (for PC/LDT1 0.300 g HC/test) and the new BETP test ("Bleed Emissions Test Procedure"), with which the tightness and the purging performance only of tank, fuel lines and carbon canister but without engine are tested (limit value for PC/LDT1 0.020 g HC per test). The durability requirement is 150,000 miles.

China

China has adopted up to and including Stage 5 the EU requirements for limiting evaporative emissions. With China 6a (7/2020) China introduced its own legislation, which combines the US procedure with WLTC elements. The limit value for the sum total of hot-soak test and 2-day diurnal test (max. temperature 38 °C, counting the higher 24-hour HC value) is 0.7 g HC.

Refueling emissions

Refueling test

In the refueling test the HC emissions when refueling are measured in order to monitor evaporation of the fuel vapors displaced during refueling (limit value Tier 2/3 / LEV II/III: 0.053 g HC per liter of pumped fuel, corresponds to 0.20 g per gallon). In the USA, this test applies to both CARB and EPA.

With China 6a (7/2020) China introduced its own legislation, which combines the US procedure with WLTC elements. The limit value is 0.05 g HC/test.

There is to date no refueling test in Europe and in Japan.

Spitback test

In the spitback test, the quantity of fuel splashed out during each refueling operation is measured. The tank must be refueled to at least 85 % of its total volume (limit value only for EPA: 1 g HC per test).

The spitback test does not have to be carried out in the refueling test; instead, a declaration that the vehicle meets the requirements of the spitback test is sufficient.

References

[1] § 29 StVZO: Untersuchung der Kraftfahrzeuge und Anhänger (Inspection of motor vehicles and trailers).

[2] ISO 10054: Internal combustion compression-ignition engines – Measurement apparatus for smoke from engines operating under steady-state conditions – Filter-type smokemeter (1998).

Diagnostics

The rise in the sheer amount of electronics in the vehicle, the use of software to control the vehicle, and the increased complexity of modern electronic systems place high demands on the diagnostic concept, monitoring in driving mode (on-board diagnostics), and workshop diagnostics.

As emission-control legislation becomes more and more stringent and continuous monitoring in driving mode is now called for, lawmakers have now acknowledged on-board diagnostics as an aid to monitoring exhaust-gas emissions, and have produced manufacturer-independent standardization. This additional system is termed the OBD system (On-Board Diagnostic system). Diagnosis of engine management systems is thus particularly important.

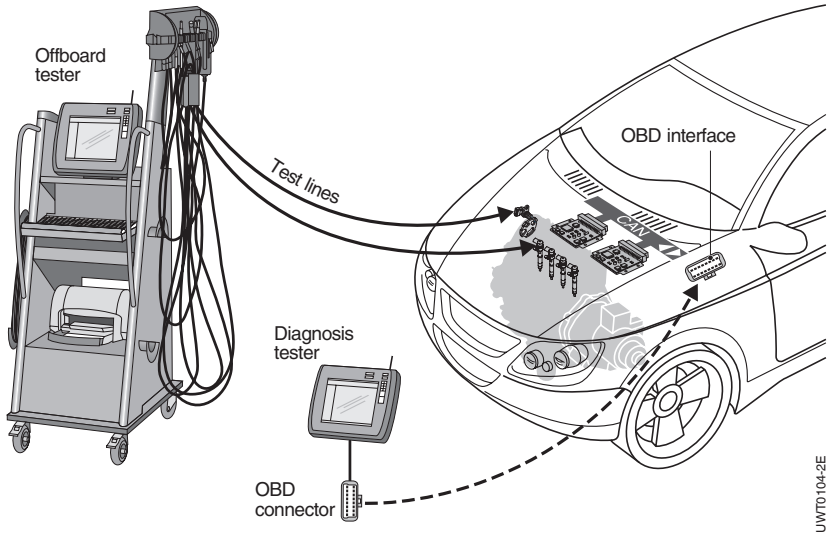
Provided for information purposes only, no liability assumed for completeness!

Monitoring in driving mode

The diagnosis integrated in the control unit is a basic feature of electronic engine management systems. Besides a self-test of the control unit, input and output signals, and control-unit intercommunication are monitored.

Monitoring algorithms check input and output signals during vehicle operation, and check the entire system and all the relevant functions for malfunctions and disturbances. Any errors or faults detected are stored in the control-unit fault memory. When the vehicle is serviced in the dealer's workshop, the stored information is exported over a serial interface. This allows troubleshooting and repairs to be carried out quickly and reliably (Figure 1).

Figure 1: Diagnostic system consisting of a diagnostic tester for reading out OBD data and an external offboard tester for further selective troubleshooting and fault localization



Monitoring input signals
The sensors, connectors and connecting lines (in the signal path) to the control unit are monitored by using the evaluated input signals. With this monitoring strategy, it is possible to detect sensor faults, short circuits to battery voltage U_B and to ground, and open circuits. The following procedures are used for this purpose:
– Monitoring sensor supply voltage (if applicable)
– Examining measures recorded for permissible value ranges (e.g. 0.5 to 4.5 V)
– Plausibility check of different physical signals (e.g. comparison of crankshaft and camshaft speeds)
– Plausibility check of a physical variable which is detected redundantly with different sensors (e.g. pedal-travel sensor)

Monitoring output signals
Actuators triggered by a control unit via output stages are monitored. The monitoring functions detect open circuits and short circuits in addition to actuator faults. The following procedures are used for this purpose: On the one hand, the electric circuit of an output signal is monitored by the output stage; the circuit is monitored for short-circuits to battery voltage U_B, to ground, and for open circuit. On the other hand, impacts on the system by the actuator are detected directly or indirectly by a function or plausibility monitor. System actuators, e.g. the exhaust-gas recirculation valve or the throttle valve, are monitored indirectly via closed-control loops (e.g. for continuous control variance), and also partly by means of position sensors (e.g. position of throttle valve).

Monitoring of internal ECU functions
Monitoring functions are implemented in control-unit hardware (e.g. "intelligent" output-stage modules) and software to ensure that the control unit functions correctly at all times. The monitoring functions check each of the control-unit components (e.g. microcontroller, flash EPROM, RAM). Many tests are performed immediately after switch-on. Further monitoring functions are repeated at regular intervals during normal operation so that component failure can also be detected during operation. Test procedures that require a high amount of computing capacity, or which cannot be performed during vehicle operation for other reasons, are carried out in the after-run after "engine off". This avoids impacting the other functions. An example of such a function is the checksum check of the flash EPROM.

Monitoring ECU communication
Communication with the other control units normally takes place over the CAN bus. Control mechanisms for error detection are incorporated in the CAN protocol so that transmission errors can be detected in the CAN chip already. The control unit also runs a variety of other tests. Since the majority of CAN messages are sent at regular intervals by the individual control units, the failure of a CAN controller in a control unit is detectable by checking these regular intervals. If redundant information is stored in the control unit, all input signals are checked against this information.

Error/fault handling
Error/fault detection
A signal path (e.g. sensor with plug connector and connecting line) is categorized as totally defective if a fault occurs over a specific length of time. The system will continue to use the last valid value until the defect is categorized. When the defect is categorized, a standby function is triggered (e.g. engine-temperature substitute value T = 90 °C).

A healing or "restored-signal recognition" feature is available during driving mode for most errors. The signal path must be detected as intact for a specific period of time for this purpose.

Error/fault storage
All faults are stored as a fault code in the non-volatile area of the data memory. The fault code also describes the fault type (e.g. short circuit, open circuit, plausibility, value range exceeded). Each fault-code input is accompanied by additional information, e.g. the operating con-

ditions (freeze frame) at the time of fault occurrence (e.g. engine speed, engine temperature).

Limp-home functions
If a fault is detected, limp-home strategies can be triggered in addition to substitute values (limp home, e.g. engine output power or speed limited). These measures are used to maintain driving safety, prevent consequential damage (e.g. catalytic converter overheating), and minimize exhaust-gas emissions.

On-board diagnostics

The engine system and components must be continuously monitored in driving mode so that compliance with the emission limits required by law can be achieved in everyday use. Therefore, starting in California, regulations were adopted to monitor exhaust-gas-related systems and components. This has standardized and expanded manufacturer-specific on-board diagnostics (OBD) with respect to the monitoring of emission-related components and systems [6].

OBD I (CARB)
In 1988 the first stage of CARB legislation (California Air Resources Board) came into force in California with OBD I. This first OBD stage requires the monitoring of emission-related electrical components (short-circuits, line breaks) and storage of the faults in the control-unit fault memory as well as a malfunction indicator lamp (MIL) that alerts the driver to detected faults. On-board means (e.g. flashing code on a diagnosis lamp) must also be in place to provide a readout of which component has malfunctioned.

OBD II (CARB)
In 1994 the second stage of diagnosis legislation was introduced in California with OBD II. OBD II became mandatory for diesel-engine cars with effect from 1996. In addition to the scope of OBD I, system functionality was now monitored (e.g. plausibility check of sensor signals).

OBD II stipulates that all emission-related systems and components must be monitored if they cause an increase in toxic exhaust-gas emissions (and thus the OBD threshold values to be exceeded) in the event of a malfunction. In addition, all components used to monitor emission-related components or which affect the diagnosis result must be monitored.

Normally, the diagnostic functions for all components and systems under surveillance must run at least once during the exhaust-gas test cycle (e.g. FTP 75, Federal Test Procedure).

The OBD II legislation furthermore prescribes standardization of the fault-memory information and access to the information (connector, communication) compliant with ISO 15031 [1] and the corresponding SAE standards (Society of Automotive Engineers), e.g. SAE J1979 [2] and SAE J1939 [3]. This permits a fault-memory readout over standardized, commercially available testers (scan tools).

OBD II expansions

From model year 2004

The law has been revised several times since OBD II was introduced. The legal requirements are generally revised by the authorities every two years ("biennial review"). Since model year 2004 it has been necessary, in addition to meeting tighter and additional functional requirements, to check the diagnostic frequency (from model year 2005) in everyday operation (In-use Monitor Performance Ratio, IUMPR).

From model year 2007 through 2013 for gasoline passenger cars

New requirements for gasoline engines are essentially the diagnosis of cylinder-individual mixture trimming (air/fuel imbalance), extended requirements with regard to diagnosis of the cold-start strategy, and permanent error/fault storage, which also applies to diesel systems.

From model year 2007 through 2013 for diesel vehicles

For diesel passenger cars and light commercial vehicles the tightened OBD emission limits are classified into three stages (up to model year 2009, model years 2010 through 2012, from model year 2013). Significantly extended functions are additionally required for the fuel-injection system, the air system, and the exhaust-gas treatment system. Thus, for example, the fuel-injection system is required to monitor the injected fuel quantity and the injection timing. The air system is required, for example, to monitor boost-pressure control and additionally the dynamic response of exhaust-gas recirculation control and boost-pressure control. In the exhaust-gas treatment system, new monitoring functions are required of the oxidation-type catalytic converter, the particulate filter, the NO_x storage catalytic converter, and the SCR dosing system (Selective Catalytic Reduction) with SCR catalytic converter. Thus, to be monitored are, for example, the regeneration frequency in the particulate filter and the dosing quantity of the NO_x reducing agent in the SCR dosing system.

One of the new requirements for diesel systems since 2009 is that, as well as the regulators, controlled functions – where they are relevant to exhaust gas – are to be monitored as well. Likewise, the monitoring of cold-start functions is subject to extended requirements.

From model year 2014/2015

Enhanced requirements have already been formulated for model year 2015 for individual components in diesel passenger cars, light commercial vehicles, and general commercial vehicles. These relate to the monitoring of the oxidation catalytic converter for "feedgas" (ratio between NO and NO_2 to operate the SCR catalytic converter), to the monitoring of the coated particulate filter for NMHC conversion (non-methane hydrocarbons), and to the monitoring of the injection system for quantity-encoded injectors. Likewise, in the course of the revision of the LEV III emission legislation some requirements pertaining to hybrid vehicles were defined more precisely with an impact on the IUMPR calculation.

From model year 2017/2023

The last revision of the OBD legislation saw among others an adaptation of the OBD threshold values to the LEV III emission legislation. The OBD II threshold value for NO_x and NMHC will be defined from LEV III as a quantity (NO_x + NMHC). The OBD threshold values as multiples of the emission limit (multipliers) were adapted in stages to the new emission categories (ULEV 50, ULEV 70, SULEV 20). For the first time an OBD threshold value was also de-

Table 1: OBD threshold values

	Gasoline passenger cars	Diesel passenger cars	Diesel commercial vehicles
CARB	– Dependent on the emission category and diagnostic requirement between 1.5 and 2.5 times the emission limit – PM OBD limit: absolute 17.5 mg/mile	– Dependent on the emission category and diagnostic requirement between 1.5 and 2.5 times the emission limit – PM OBD limit: absolute 17.5 mg/mile – But from 2007 to 2013 introduction of stricter limits in three stages: e.g. for particulate filter 2007 – 2009 5 x limit 2010 – 2012 4 x limit from 2013 1.75 x limit	2010–2012: CO: 2.5 × limit NMHC: 2.5 × limit NO_x: +0.4/0.6 g/bhp-hr [2] PM: +0.06/0.07 g/bhp-hr From 2013: CO: 2.0 × limit NMHC: 2.0 × limit NO_x: +0.2/0.4 g/bhp-hr [2] PM: +0.02/0.03 g/bhp-hr Transition phase for some monitors up to 2016
EPA (US Federal)	see CARB CARB certificates with corresponding limits are recognized by EPA	see CARB CARB certificates with corresponding limits are recognized by EPA	2010–2012: CO: 2.5 × limit NMHC: 2.5 × limit NO_x: +0.6/0.8 g/bhp-hr [2] PM: +0.04/0.05 g/bhp-hr From 2013: CO: 2.0 × limit NMHC: 2.0 × limit NO_x: +0.3/0.5 g/bhp-hr [2] PM: +0.04/0.05 g/bhp-hr
EOBD	Euro 5 (09/2009): CO: 1,900 mg/km NMHC: 250 mg/km NO_x: 300 mg/km PM: 50 mg/km [1] Euro 6-1 (09/2014): CO: 1,900 mg/km NMHC: 170 mg/km NO_x: 150 mg/km PM: 25 mg/km [1] Euro 6-2 (09/2017): CO: 1,900 mg/km NMHC: 170 mg/km NO_x: 90 mg/km PM: 12 mg/km [1]	Euro 5 (09/2009): CO: 1,900 mg/km NMHC: 320 mg/km NO_x: 540 mg/km PM: 50 mg/km Euro 6 interim (09/2009): CO: 1,900 mg/km NMHC: 320 mg/km NO_x: 240 mg/km PM: 50 mg/km Euro 6-1 (09/2014): CO: 1,750 mg/km NMHC: 290 mg/km NO_x: 180 mg/km PM: 25 mg/km Euro 6-2 (09/2017): CO: 1,750 mg/km NMHC: 290 mg/km NO_x: 140 mg/km PM: 12 mg/km	Euro IV (10/2005)/ Euro V (10/2008): NO_x: 7.0 g/kWh PM: 0.1 g/kWh NO_x control-system monitoring (since 11/2006): NO_x emission limit + 1.5 g/kWh Euro IV: (3.5+1.5) g/kWh Euro V: (2.0+1.5) g/kWh Euro VI-A (2013): NO_x: 1.5 g/kWh PM: 0.025 g/kWh (CI engine) Functional alternative for DPF monitor NO_x control system: SCR reagent NO_x: 0.9 g/kWh Euro VI-B (09/2014): SI engine Like Euro VI-A, but threshold value for CO: 7.5 g/kWh Euro VI-C (2016): NO_x: 1.2 g/kWh PM: 0.025 g/kWh (CI engine) CO: 7.5 g/kWh (SI engine) NO_x control system: SCR reagent NO_x: 0.46 g/kWh

[1] For gasoline direct injection.

[2] g/bhp-hr: grams per “brake horse power” × “hour” (brake horse power corresponds to the German unit PS).

fined for the particulate matter for gasoline engines from model year 2019 at a fixed value of 17.5 mg per mile. Further new requirements are, for example, the output of characteristics for assessing the use and activation of "Active off-cycle technologies" and fuel-consumption-specific quantities from model year 2019, improved monitoring of the crankcase-breather lines from model year 2023, and specification of a multitude of diagnostic requirements for components of hybrid vehicles.

Considerations are being given to extending the OBD requirements to CO_2 monitoring for future legislation.

Scope of application
The previously presented OBD regulations for the CARB and apply to all passenger vehicles with up to 12 seats as well as small commercial vehicles up to 14,000 lbs (6.35 t).

The current CARB OBD II legislation for California is at present also in force in some other US states. Furthermore, other US states are planning to adopt this legislation in future.

EPA OBD
Laws enforced by the EPA (Environmental Protection Agency) have been in force since 1994 in those US states which have not adopted the CARB legislation. The requirements of this diagnosis are essentially equivalent to the CARB legislation (OBD II). Within the framework of revising the Tier 3 emission legislation, from model year 2017 the EPA OBD requirements were adapted to the CARB OBD requirements. A CARB certificate is already recognized now by the EPA.

EOBD (European OBD)
On-board diagnostics attuned to European conditions is termed EOBD. EOBD has applied since January 2000 to passenger cars and light commercial vehicles with gasoline engines. The regulation has been valid for passenger cars and light commercial vehicles equipped with diesel engines since 2003, and for heavy commercial vehicles since 2005 (see OBD requirements for heavy commercial vehicles).

In 2007 and 2008 new EOBD requirements were adopted for gasoline and diesel passenger cars within the framework of the Euro 5 and Euro 6 emission and OBD legislation (Euro 5 emission level from September 2009; Euro 6 from September 2014).

A general new requirement for gasoline and diesel passenger cars is for checking the diagnostic frequency in everyday operation (In-Use Performance Ratio, IUPR) in accordance with the CARB OBD legislation (In-use Monitor Performance Ratio, IUMPR) as from Euro 5+ (September 2011).

EOBD Euro 5 and Euro 5+ requirements for diesel and gasoline engines
For gasoline engines, the introduction of Euro 5 as from September 2009 involved primarily a reduction in the OBD threshold values. In addition to a particulate-matter OBD threshold value (for direct-injection engines only), an NHMC OBD threshold value was introduced (non-methane hydrocarbons, instead of the previous HC). Direct functional OBD requirements result in the monitoring of the three-way catalytic converter for NMHC. Since September 2011 the Euro 5+ level has applied with unchanged OBD threshold values compared with Euro 5. A significant functional requirement with regard to EOBD is the additional monitoring of the three-way catalytic converter for NO_x.

For diesel passenger-car engines, Euro 5 involved a reduction of the OBD threshold values for particulate matter, CO and NO_x. In addition, there are extended requirements with regard to the monitoring of the exhaust-gas recirculation system (cooler) and above all with regard to the exhaust-gas treatment components. Here, monitoring of the SCR DeNOx system (dosing system and catalytic converter) is subject to very stringent requirements. Functional monitoring of the particulate filter is mandatory, irrespective of the untreated emissions.

EOBD Euro 6 requirements for diesel and gasoline vehicles
With Euro 6-1 from September 2014 and Euro 6-2 from September 2017 a further two-stage reduction of some OBD threshold values has been decided on (see Table 1). Furthermore, diesel systems are governed by tighter regulations for monitoring the oxidation catalytic converter and the NO_x exhaust-gas treatment system (NO_x storage catalytic converter or SCR catalytic converter with dosing system).

From September 2017 the NEDC is being replaced by the WLTC with Euro 6d-temp for the Type 1 emission test [8]. Here the emission limits as well as the OBD threshold values have not been adapted, but rather adopted unchanged. With regard to the OBD test the vehicle manufacturer can choose whether the die OBD threshold-value check is carried out on the basis of the NEDC or the WLTC. This right to choose will only exist until the end of 2021. From this point on, the OBD threshold values will exclusively be checked on the basis of the WLTC.

China OBD
In December 2016 the MEP (Ministry of Environmental Protection of the People's Republic of China) published a new law with significantly tightened emission and OBD requirements will come into force with regard to emissions in two stages (CN6a from 07/2020 and CN6b from 07/2023). The OBD requirements from 07/2020 will also apply unchanged to stage CN6b. While the previous Chinese legislation was geared very closely towards the European standard, the new CN6 law combines elements of the EU and US legislation and some country-specific new requirements. With regard to OBD, the requirements are based to the greatest possible extent on US requirements of the legislation from 2013, where some requirements have been removed or simplified and others have been enhanced. While the actual OBD requirements are geared towards the US standard, the European OBD threshold values for Euro 6-2 and the European test cycle WLTC have on the other hand been adopted.

Other countries
Some other countries have adopted different stages of the EU or the US OBD legislation (Russia, South Korea, India, Brazil, Australia).

OBD system requirements
The engine control unit must use suitable measures to monitor all on-board systems and components whose malfunction may cause a deterioration in exhaust-gas test specifications stipulated by law. A malfunction must be displayed to the driver by means of the malfunction indicator lamp (MIL) if a fault results in an excess in OBD threshold values.

OBD threshold values
US OBD II (CARB and EPA) prescribes thresholds that are defined based on emission limits. Accordingly, there are different permissible OBD threshold values for the various exhaust-gas categories that are applied during vehicle certification (e.g. LEV, ULEV, SULEV). The European EOBD regulations are based on absolute threshold values (Table 1).

Functional requirements
All exhaust-gas-related systems and components must, within the framework of On-Board Diagnostics (OBD) required by law, be monitored for malfunctions and for exceeded exhaust-gas threshold values.

Legislation demands the monitoring of electrical functions (short-circuit, line breaks), a plausibility check for sensors, and a function monitoring for actuators.

The pollutant concentration expected as the result of a component failure (can be measured in the exhaust-gas cycle) and the monitoring mode partly required by law determine the type of diagnostics. A simple functional test (black/white test) only checks system or component operability (e.g. swirl control valve opens or closes). The extensive functional test provides more detailed information about system operability and also identifies if necessary the quantitative influence of

defective components on emissions. As a result, the limits of adaptation must be monitored when monitoring adaptive fuel-injection functions (e.g. zero delivery calibration for a diesel engine, λ adaptation for a gasoline engine).

Diagnostic complexity has constantly increased as emission-control legislation has evolved.

Malfunction indicator lamp
The malfunction indicator lamp (MIL), also called the warning lamp, informs the driver that a component has malfunctioned. When a malfunction is detected, the CARB and the EPA stipulate that it must light up no later than after one driving cycle of its occurrence. In the area where EOBD applies, it must light up no later than the third driving schedule after the fault was detected.

If the malfunction disappears (e.g. a loose contact), the malfunction remains entered in the fault memory for 40 trips ("warm-up cycles"). The malfunction indicator lamp goes out after three fault-free driving schedules. The malfunction indicator lamp flashes for faults in the gasoline system if such a fault could cause damage to the catalytic converter (e.g. combustion misses).

Communication with the scan tool
The OBD legislation prescribes standardization of the fault-memory information and access to the information (connector, communication interface) compliant with the standard ISO 15031 and the corresponding SAE standards (e.g. SAE J1979, [2]). This permits a fault-memory readout over standardized, commercially available testers (scan tools).

Only diagnostics via CAN (ISO 15765 [4]) has been permitted for CARB since 2008 and for the EU since 2014.

Vehicle repair
Any workshop can use a scan tool to read out emission-related fault information from the control unit. This permits even non-franchised workshops to carry out repairs.

Manufacturers are obliged to provide the required tools and information (repair manuals on the Internet), for a suitable fee, to ensure that repairs can be carried out with the correct degree of expertise.

Switch-on conditions
The diagnostic functions are only executed if the physical switch-on conditions are fulfilled. These include, for example, torque thresholds, engine-temperature thresholds, and engine-speed thresholds or limits.

Inhibit conditions
Diagnostic functions and engine functions cannot always operate simultaneously. There are inhibit conditions that prohibit the performance of certain functions. For instance, tank ventilation (with evaporative-emissions control system) in the gasoline engine cannot function while catalytic-converter diagnosis is in operation. In the diesel engine the hot-film air-mass meter can only be monitored satisfactorily if the exhaust-gas recirculation valve is closed.

Temporary interruption of diagnostic functions
Diagnostic capabilities may only be disabled under certain conditions in order to prevent false diagnosis. Examples of such conditions are high elevation (low air pressure), during engine starting low ambient temperature, or low battery voltage.

Readiness codes
When the fault memory is checked, it is important to know that the diagnostic functions have run at least once. This can be checked by reading out the readiness codes over the diagnostics interface. These readiness codes are set for the most important monitored components on completion of the relevant diagnoses required by law.

Diagnostic System Management
The diagnostic capabilities for all components and systems checked must regularly run in driving mode, but also at least once during the exhaust-gas test cycle (e.g. FTP 75, NEDC). The Diagnostic System Management (DSM) can dynamically change the sequence for running the diagnostic functions depending on the driving condition. The objective here is to run the diagnostic functions frequently in everyday vehicle operation. Diagnostic System Management consists of the following components:
– Diagnostic Fault Path Management for storing fault states and associated ambient conditions (freeze frames),
– Diagnostic Function Scheduler for coordinating the engine and diagnostic functions,
– Diagnostic Validator for deciding centrally when faults are detected whether they are the cause or a consequential fault. As well as central validation there are also systems with decentralized validation, i.e. validation is performed in the diagnostic function.

Vehicle recall
If vehicles fail to comply with OBD requirements by law, the authorities may demand the vehicle manufacturer to start a recall at their own cost.

OBD functions

Overview
Whereas EOBD only contains detailed monitoring specifications for individual components, the specific requirements in CARB OBD II are much more detailed. The list below shows the current state of the most significant CARB requirements (as at 2017) for gasoline-engined and diesel-engined vehicles. The requirements that are also described in detail in the EOBD legislation are marked by (E).

Gasoline and diesel system:
– Exhaust-gas recirculation system (E),
– Cold-starting emission-control system,
– Crankcase ventilation,
– Combustion misses/misfiring (E, for gasoline system only),
– Fuel system,
– Variable valve timing,
– Exhaust-gas sensors (λ oxygen sensors (E), NO_x sensors (E), particulate sensor),
– Engine cooling system,
– Air-conditioning system (in the event of influence on emissions or on OBD)
– Other emission-related components and systems (E),
– In-use Monitor Performance Ratio (IUMPR) for checking the frequency of diagnostic functions in everyday operation (E),
– Requirements which apply to diesel/(gasoline) engines must, if the same technologies are used for the gasoline/(diesel) engine, be evaluated according to the diesel/(gasoline) requirements and the diagnostic concept of the authority/agency presented.

Gasoline system only:
– Secondary-air injection,
– Three-way catalytic converter (E), heated catalytic converter,
– Tank-leak diagnosis, with (E) at least electrical testing of the canister-purge valve,
– Direct ozone-reduction system,
– Cylinder-specific λ irregularity.

Diesel system only:
– Oxidation-type catalytic converter (E),
– SCR DeNOx system (E),
– NO_x storage catalytic converter (E),
– Particulate filter (E),
– Fuel-injection system (rail-pressure control, injected fuel quantity, and injection timing),
– Cooler for exhaust-gas recirculation (E),
– Boost-pressure control,
– Intercooler.

Other emission-related components and systems refer to components and systems not mentioned in this list that, if they malfunction, may cause the exhaust-gas emissions to be increased (CARB-OBD II), the OBD threshold values to be exceeded (CARB OBD II and EOBD), or the diagnostic system to be negatively influenced (e.g. by inhibiting other diagnostic functions). Minimum values must be maintained with regard to the frequency of diagnostic functions.

Examples of OBD functions

Catalytic-converter diagnosis

Gasoline system

This diagnostic function monitors the conversion efficiency of the three-way catalytic converter. This is measured by the catalytic converter's oxygen retention capability. Monitoring is performed by observing the signals from the λ oxygen sensors in reaction to a specific alteration of the setpoint value of the λ closed-loop control.

Additionally, the NO_x storage capacity (catalytic-converter quality factor) must be assessed for the NO_x storage catalytic converter. For this purpose, the actual NO_x accumulator content resulting from consumption of the reduction agent during regeneration of the catalytic converter is compared with an expected value.

Diesel system

In the diesel system, carbon monoxide (CO) and unburned hydrocarbons (HC) are oxidized in the oxidation-type catalytic converter (pollutant minimization). Diagnostic functions are used to monitor the operation of the oxidation-type catalytic converter on the basis of the temperature differential before and after the catalytic converter (exothermy).

The storage and regeneration capacity of the NO_x storage catalytic converter is monitored. The monitoring functions run based on loading and regeneration models, and the measured regeneration duration. This requires the use of λ or NO_x sensors.

The SCR DeNOx catalytic converter is monitored using efficiency diagnostics. NO_x sensors located before and after the catalytic converter are required for this purpose. The components of the dosing system and the quantity and dosing of the reducing agent are monitored separately.

Tank-leak diagnosis

Gasoline system

Tank-leak diagnosis detects evaporation from the fuel system that may cause an increase in HC values, in particular. EOBD is limited to simply testing the electrical control circuit of the tank-pressure sensor and the canister-purge valve (evaporative-emissions control system). In the USA, on the other hand, it must be possible to detect leaks in the fuel system. There are two different methods of doing this.

The low-pressure method observes the tank pressure and first tests its operability by deliberately actuating the tank ventilation and carbon-canister check valves. A conclusion can then be drawn on the leak size using the time curve of the tank pressure – again by deliberately actuating the valves.

The overpressure method uses a diagnosis module with an integrated electrically powered vane pump that can be used to pump up the tank system. The flow from the pump is high when the tank is hermetically sealed. A conclusion can then be drawn on leak size by evaluating the flow from the pump.

Particulate-filter diagnosis
Diesel system
The diesel particulate filter is currently monitored for the most part for filter breakage, removal, or blockage. A differential-pressure sensor is used to measure the pressure differential (exhaust-gas backpressure downstream and upstream of the filter) at a specific volumetric flow. The measured value can be used to verify whether the filter is defective. An extended function uses load models to monitor the efficiency of the particulate filter.

Since model year 2010 the regeneration frequency has also had to be monitored. Since model year 2013 a particulate sensor has been used to monitor the particulate filter in response to tightened OBD requirements in the USA. The particulate sensor (from Bosch) operates according to the "collection principle", i.e. the soot collected over a specific distance driven is evaluated using a model for a threshold-value filter. If the collected soot matter, as a function of different parameters, exceeds a certain threshold, the particulate filter is detected as faulty. Combined particulate-filter faults (e.g. broken and melted filters) can also be detected with the particulate sensor.

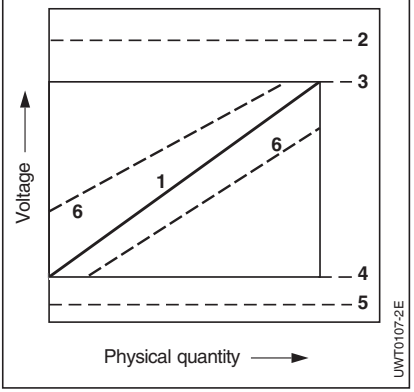
Exhaust-gas recirculation system diagnosis
Diesel system
In the exhaust-gas recirculation system (EGR) the regulator, the exhaust-gas recirculation valve, the exhaust-gas cooler, and other individual components are monitored.

Functional system monitoring is performed by air-mass regulators and position controllers. They check for permanent control variances. An excessively high or low EGR throughflow must be detected. System response ("slow response") must also be monitored.

The exhaust-gas recirculation valve itself is monitored for its electrical and functional operability.

The EGR cooler is monitored by means of additional temperature measurement after the cooler and with model values. This allows the cooler efficiency to be calculated.

Figure 2: Sensor monitoring
1 Sensor curve,
2 Upper threshold for "Signal Range Check",
3 Upper threshold for "Out of Range Check",
4 Lower threshold for "Out of Range Check",
5 Lower threshold for "Signal Range Check",
6 Plausibility check "Rationality Check".



Comprehensive components
On-board diagnostics requires that all sensors (e.g. air-mass meter, speed sensor, temperature sensors) and actuators (e.g. throttle valve, high-pressure pump, glow plugs) having either an impact on emissions or are being used to monitor other components or systems (and consequently may if necessary disable other diagnoses) have to be monitored.

Sensors monitor the following faults (Figure 2):
– Electrical faults, i.e. short-circuits and line breaks ("Signal Range Check"),
– Range faults ("Out of Range Check"), i.e. undercutting or exceeding voltage limits set by the sensors' physical measurement range,
– Plausibility faults ("Rationality Check"); these are faults that are inherent in the components themselves (e.g. drift), or which may be caused by shunts, for instance. Monitoring is carried out by a plausibility check on the sensor signals, either by using a model or directly by other sensors.

Actuators must be monitored for electrical faults and – if technically possible – also for function. Functional monitoring means that, when a control command (setpoint value) is given, it is monitored by observing or measuring (e.g. by a position sensor) the system reaction (actual value) in a suitable way by using information from the system.

The actuators to be monitored include all output stages, the throttle valve, the exhaust-gas recirculation valve, the variable turbine geometry of the exhaust-gas turbocharger, the swirl flap, the injectors, the glow plugs (for diesel systems), the tank-ventilation system (for gasoline systems), and the active-charcoal check valve (for gasoline systems).

OBD requirements for heavy commercial vehicles

Europe
For commercial vehicles, the first stage of On-Board Diagnostics was introduced in the EU (EOBD) together with Euro IV (10/2005), the second stage together with Euro V (10/2008). A new OBD regulation came into force together with Euro VI in 2013.

Monitoring requirements in Stage 1
– Fuel-injection system: monitoring for electrical faults and for total failure,
– Engine components: monitoring of emission-related components for compliance with the OBD threshold value,
– Exhaust-gas treatment systems: monitoring for serious faults.

Additional requirements in Stage 2
– Exhaust-gas treatment systems: monitoring for compliance with the OBD threshold value.

Additional requirements
Since November 2006 it has been required to monitor the NO_x control systems for correct operation. The systems are monitored for their own emission limits, which are more stringent than the OBD threshold values.

SCR system
The aim is to ensure that the systems are supplied with the correct reagent (urea/water solution, the customary brand name is AdBlue). The availability of the reagent must be monitored by way of the tank fill level. To check the correct quality, it is necessary to monitor the NO_x emissions either with an exhaust-gas sensor or alternatively via a quality sensor. In the latter case, it is also necessary to monitor for correct reagent consumption.

Exhaust-gas recirculation system
The exhaust-gas recirculation system is monitored for correctly recirculated exhaust-gas mass flow and for deactivation of exhaust-gas recirculation.

NO_x storage catalytic converters
Exhaust-gas sensors are used to monitor NO_x emission.

Monitoring of NO_x control systems
Faults in NO_x control systems must be permanently stored (i.e. non-erasable) for 400 days (9,600 hours). Engine power must be throttled if the NO_x OBD threshold value is exceeded or if the urea tank is empty.

Euro VI
The OBD part of the Euro VI regulation is based on the Global Technical Regulation (GTR) "World Wide Harmonized OBD" (WWH OBD). In terms of structure this WWH OBD GTR corresponds to the Californian OBD legislation (passenger cars and commercial vehicles).

WWH OBD leaves opens which monitoring operations are actually selected in a national regulation (here Euro VI) which implements WWH OBD.

In addition, emission limits and OBD threshold values as well as the choice of test cycles are established by way of the national regulations.

Special aspects of WWH OBD involve the introduction of new fault storage and new scan-tool communication (ISO 27145 [5]).

Faults must be classified according to their severity with regard to emission deterioration. Emission-related faults can be distinguished by means of malfunction-indicator-lamp behavior and scan-tool communication. The following different categories are applied:
– A: Emissions above the OBD threshold value,
– B1: Emissions above or below the OBD threshold value,
– B2: Emissions below the OBD threshold value, but above the emission limit,
– C: Influence on emissions below the emission limit.

According to this principle all emission-related faults are output, even those which exert a very small influence.

Data for Euro VI
– Drastic reduction of the emission limits and OBD threshold values for NO_x and particulate matter compared with Euro V,
– Emission limits to be introduced for NH_3 and the number of particulates,
– Use of the harmonized WHSC and WHTC test cycles,
– The OBD demonstration is performed with a double WHTC hot start part,
– Introduction of a facility for checking system conformity with regard to emissions in the field by means of random measurements with portable emission measurement systems (PEMS),
– Checking of the diagnostic frequency of OBD monitoring in everyday operation (In-Use Monitoring, IUMPR).

Euro VI A
– For new type approvals mandatory since 31 December 2012,
– Valid up to 31 August 2015,
– Strict OBD threshold values for NO_x and particulate matter (only for engines with auto-ignition). For particulate-filter monitoring a functional, non-emission-correlated diagnosis is possible as an alternative to OBD threshold-value diagnosis.

Euro VI B
– For new type approval mandatory as from 1 September 2014,
– Valid up to 31 December 2016,
– Affects only engines with externally supplied ignition,
– Introduction of an OBD threshold value for CO.

Euro VI C
– For new type approval mandatory as from 31 December 2015,
– Changes in respect of Euro VI A and Euro VI B: Tightened NO_x OBD threshold value and tightening of the NO_x control-system requirements for SCR reagent quality and consumption monitoring. Monitoring is performed with regard to the long-term drift behavior of the fuel injectors. Monitoring of the OBD diagnostic-frequency rate is mandatory.

Euro VI D and Euro VI E
– For new type approval mandatory as from 1 September 2018 or 1 January 2021,
– No change with regard to OBD.

Required diagnoses from WWH OBD
These diagnoses are compulsory for particulate filter, SCR catalytic converter, NO_x storage catalytic converter, oxidation catalytic converter, exhaust-gas recirculation, fuel-injection system, charge-air pressure system, variable valve control, cooling system, exhaust-gas sensors, idle control system, and components.

Required diagnoses outside the WWH OBD scope
For the particulate filter, the exhaust-gas recirculation system and the charge-air pressure control system no exceptions are permitted in monitoring for specific diagnoses. The relevant faults must not be defined as Class C.

Furthermore, as from Euro VI C the monitoring of possible component-damaging effects of a long-term drift of fuel injectors is required.

Fundamentally the definition of the non-emission-correlated "performance monitors" in WWH OBD has been changed. In Euro VI these diagnoses must be demonstrated with emission correlation for the first certification of an engine from an engine family.

Gas engines
Gas engines are covered by specific monitoring requirements with regard to compliance with the λ setpoint value, to NO_x and CO conversion of the three-way catalytic converter, and to the λ sensor. Furthermore, catalytic-converter-damaging combustion-miss detection is required.

Requirements with regard to NO_x control systems
For SCR systems, monitoring of the reagent tank fill level, of the reagent quality, of reagent consumption, and of dosage interruption is required.

For exhaust-gas recirculation systems, monitoring of the exhaust-gas recirculation valve is required.

Furthermore, all the NO_x control systems must be monitored for system deactivation by tampering.

Detected faults in the NO_x control system result in a reduction in stages of vehicle drivability. Torque limitation as the first stage is followed by vehicle-speed limitation to crawling speed as the second stage.

USA
CARB, model year 2007
In California "Engine Manufacturer Diagnostics" (EMD) has been called on for heavy commercial vehicles since model year 2007. This can be viewed as a precursor to an OBD regulation. EMD calls for the monitoring of all components and of exhaust-gas recirculation.

The requirements are not reflected in a separate emission limit. Furthermore, no standardized scan-tool communication is required.

Model year 2010 and following
With model year 2010 an OBD system as for passenger-car OBD II was introduced. The technical requirements are at the same level as the respective requirements for passenger cars.

Differences ensue due to the fact that commercial vehicles are subject to engine certification. All the emission limits and OBD threshold values apply here to engine cycles. The absolutely applicable values are scaled with the work performed in the cycle.

In contrast to passenger-car LEV III, a new emission regulation (introducing an NO_x and NMHC aggregate limit value) is not planned for commercial vehicles. The OBD threshold values for NO_x and NMHC thus remain unchanged, i.e. separate, for commercial vehicles.

Time frame for introducing OBD requirements
Model year 2010
A performance variant of a manufacturer's top-selling engine family must be equipped with an OBD system. The other performance variants of this engine family are covered by a simplified certification procedure.

Model year 2013
An engine family of a manufacturer must be equipped in all the performance variants with an OBD system. Furthermore, an OBS system is required in each case for a performance variant in every engine family. The other performance variants of these engine families are covered by a simplified certification procedure.

Model year 2016
All the engine families of a manufacturer must have an OBD system in all the performance variants.

Model year 2018
Engines powered by alternative fuels (e.g. gas) are subject to the OBD requirements.

Japan
Japan has had its own OBD regulation for commercial vehicles in force since 2004. The requirements are comparable in terms of content with EMD in California for model year 2007.

China
China VI has applied in China since 2020; this is based on Euro VI and additionally imposes some further requirements that also relate to OBD.
- Additional demonstration of the OBD diagnoses at the vehicle on a special commercial-vehicle chassis-dynamometer test cycle C-WTVC,
- Monitoring of the injected fuel quantity with regard to the OBD threshold values,
- Remote transmission of live operating data which are determined via the OBD system by way of remote data transmission to a server,
- So-called “Permanent DTC” fault storage similar to CARB OBD, adapted to WWH OBD fault storage,
- Special temperature monitoring for vanadium catalytic converters,
- Use of a warning and prompt system, as there is for Euro VI for NO_x control-system requirements, for OBD faults too. This relates to the OBD faults for a loss of efficiency of the most important exhaust-gas treatment systems DPF (diesel particulate filter), SCR catalytic converter (selective catalytic reduction), NO_x storage catalytic converter, and three-way catalytic converter,
- Specifically for gas engines: monitoring of crankcase ventilation.

Other countries
Further countries have in the meantime introduced OBD for commercial vehicles. These include India, Korea, Australia, Brazil, and Russia. These countries have adopted EU regulations in this regard (Euro IV, Euro V or Euro VI (Korea)).

On-board diagnostics for motorcycles

While there have hitherto been no OBD requirements for motorcycles in the USA, OBD is being introduced in three stages in the EU. Emission stage Euro 4 as from 2016 saw the introduction of OBD I, which requires the monitoring of electrical faults for emission-relevant sensors and actuators. Which diagnoses of electrical faults are required is dependent on the component. The additional requirements such as fault detection and fault storage, MIL activation, scan-tool communication, etc. conform to the EOBD passenger-car requirements.

With emission stage Euro 5 as from 2020 OBD I will be extended to include plausibility checks for sensors and function checks for actuators. The diagnostic range is dependent on the component. It will furthermore be necessary to monitor for faults which lead to a reduction of the maximum torque by more than 10 %. At the same time OBD II requirements as for EOBD apply as well. Catalytic-converter diagnosis is excluded. The detection of misfires on motorcycles is covered by engine-speed load ranges. The functional IUPR requirements must be satisfied, but without prescribed minimum ratios. From 2024 reduced OBD threshold values will reply; furthermore, catalytic-converter diagnosis is prescribed and a minimum ratio of 0.1 will apply to all diagnoses.

The first OBD I stage was carried over to UN GTR 18, which alternatively also provides for reduced scopes. The carry-over of OBD I stage 2 and of OBD II in Amendment 1 to UN GTR 18 took place in 2020. India introduced EU OBD I from 4/2020 and is expected to introduce EU OBD II from 4/2023. Discussions are being held as to whether the OBD II stage is, as in the EU, to be split into two stages.

Japan introduced its own OBD I with a similar scope to EU OBD I in 2016. A further stage OBD II is to be based on the previously mentioned UN GTR OBD II.

China has its own simplified variant of the EU OBD I legislation, valid from 7/2018.

ECU diagnostics and Service Information System

Function
The function of diagnostics in the workshop is to locate quickly and reliably the smallest replaceable defective unit. In modern vehicles it is essential to use a generally PC-based diagnostic tester. Diagnostics in the workshop utilizes the results of diagnostics conducting in driving mode (fault-memory entries), employing special workshop diagnostic modules in the vehicle ECU or diagnostic tester and additional test and measuring equipment. These diagnostic options are integrated in the diagnostic tester in prompted troubleshooting.

Prompted troubleshooting
The main element of workshop diagnostics is the prompted troubleshooting procedure. The workshop employee is guided, starting out from the fault pattern or from the fault-memory entry, through the relevant diagnostic steps. The choice and order of the diagnostic steps is dynamic and depends on the result of the previous diagnostic step. The diagnostic results contain the use of test equipment, additional sensors or workshop diagnostic modules.

Symptoms
Faulty vehicle behavior can be either perceived directly by the driver or documented by way of a fault-memory entry. At the beginning of fault diagnosis the workshop employee must identify the existing symptom as the starting point for prompted troubleshooting.

Fault-memory entries
All the faults that occur during driving are stored in the fault memory together with defined ambient conditions prevailing at the time of their occurrence, and can be read out via an interface protocol. This protocol is based on one of the established standards and is generally extended to include manufacturer-specific components. The fault memory can also be deleted with the diagnostic tester.

Additional test equipment and sensors
The diagnostic options in the workshop are expanded by using additional sensors (e.g. clamp-on current pickup, clamp-on pressure pickup) or test equipment (e.g. Bosch vehicle-system analyzer). In the event of a fault detected in the workshop, the equipment is adapted to the vehicle. The measurement results are generally evaluated by the diagnostic tester.

Workshop diagnostic modules
All the workshop diagnostic modules can be used only when a diagnostic tester is connected and generally only when the vehicle is stationary. The operating conditions are monitored in the ECU. These diagnostic modules integrated in the ECU, after being started by the diagnostic tester, run completely autarkically in the ECU and on completion signal the result back to the diagnostic tester. ECU-based workshop diagnostic modules differ from simple actuator tests with acoustic feedback in that they can place the vehicle to be diagnosed in the workshop into as a rule no-load operating points, excite actuators depending on the boundary conditions set, and independently evaluate the result via sensor values with an evaluator logic circuit. Examples of such modules are the high-pressure test as system test for the diesel injection system (Figure 1) and the run-up test as component test for the diesel injectors (Figure 2).

Tester-based diagnostic modules
The functional sequence and evaluation are executed in the case of tester-based diagnostic modules in the diagnostic tester, where the measurement data used for evaluation are determined with the aid of the ECU by sensors in the vehicle or by additional test sensors.

The performance of the tester-based diagnostic modules is dependent on the released interfaces in the engine ECU and the data transfer between engine ECU and tester. On the other hand, high flexibility

Figure 1: High-pressure test for the diesel injection system

Detection of leaks in the high-pressure system and efficiency of high-pressure generation.
Steps:
Start test – Increase setpoint pressure – Measure pressure build-up time – Reduce setpoint pressure – Measure pressure reduction time – Variation of engine speed and pressure – Measure pressure reduction time with engine stopped at end of test – Diagnostic results.

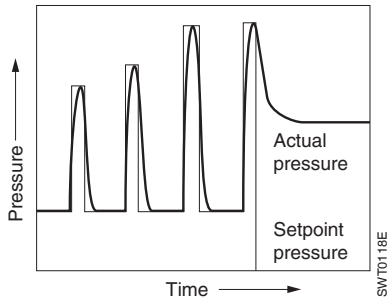
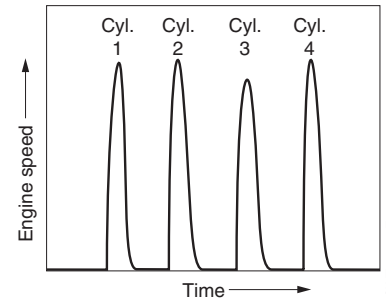


Figure 2: Run-up test for the diesel injectors

Detection of deviations of injected fuel quantity of individual injectors.
Steps:
Start test – Cutout of individual cylinder – Quantity jump – Measure maximum speed – Repeat with further cylinders – Diagnostic result.



is also possible after series introduction since new diagnostic modules can also be used thanks to a simple update of the tester software. Selective diagnostics is therefore possible when field problems occur.

Further contents of fault-finding instructions

The fault-finding instructions also include function descriptions, electrical terminal diagrams, installation positions for components, read-outs of actual values, and actuator diagnostics.

References

[1] ISO 15031: Road vehicles – Communication between vehicle and external equipment for emissions-related diagnostics (2011).
[2] SAE J 1979: E/E Diagnostic Test Modes (2012).
[3] SAE J1939: Serial Control and Communications Heavy Duty Vehicle Network – Top Level Document (2012).
[4] ISO 15765: Road vehicles – Diagnostics over Controller Area Network (2011).
[5] ISO 27145: Road vehicles – Implementation of World-Wide Harmonized On-Board Diagnostics (WWH-OBD) communication requirements (2012).
[6] OBD II regulation, section 1968.2 of title 13, California Code of Regulations, different approved OAL versions.
[7] UN/ECE Regulation No. 83, Revision 5: Uniform provisions concerning the approval of vehicles with regard to the emission of pollutants according to engine fuel requirements.
[8] (WLTP) Regulations (EU) 2017/1151 and 2017/1347.

Chassis systems

Overview

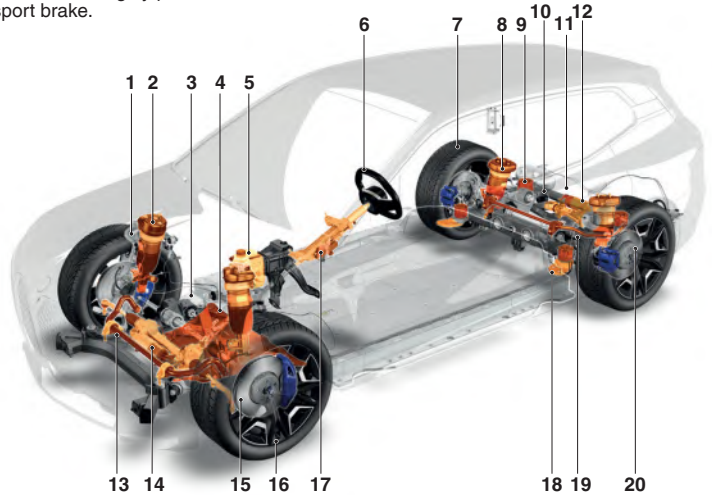
Definition

Aside from the drive (engine, transmission) and the vehicle body (body, interior), the chassis constitutes one of the classic, property-determining main assemblies of a motor vehicle. As the link between the vehicle body and the road, the chassis is crucial to generating and transmitting the horizontal and vertical forces between the tires and the road which make it possible to propel, brake, and steer a vehicle. With the exception of the aerodynamic influences all the external forces and moments are introduced to the vehicle via the tire contact area (tire contact patch, footprint). During driving it is imperative that contact between the tire and the roadway is not lost. The higher the driving speed, the more exacting the demands on the transmission of force to the tire contact patches. The chassis and its properties therefore have a significant impact on a vehicle's driving dynamics (handling), ride comfort, and driving safety.

Because of the many and sometimes contrary requirements imposed on a vehicle between the conflicting priorities of driving dynamics, ride comfort, and driving safety, the design and tuning of a chassis are subject to high levels of complexity.

Bild 1: Fahrwerkskomponenten eines Elektrofahrzeugs (Bild: BMW iX)

1 Double-wishbone front axle, 2 Two-axle pneumatic suspension with electronically controlled vibration damper, 3 Actuator-contiguous wheel-slip limitation for two driving engines,
4 Front-axle support stiffening plate, 5 Integrated braking system, 6 Polygonal steering wheel,
7 Rolling-resistance- and acoustics-optimized 22" tires with foam insert,
8 Two-axle pneumatic suspension with electronically controlled vibration damper,
9 Bearing, rear-axle support, 10 Functional five-link rear axle with double acoustic isolation of drive unit, 11 Actuator-contiguous wheel-slip limitation for two driving engines,
12 Integral active steering, 13 Anti-roll bar with highly pretensioned anti-roll bar mounts,
14 Steering with variable rack geometry and acoustic isolation, 15 17" sport brake,
16 Air Performance wheels, 17 Electric steering-column adjustment,
18 Compression strut connecting rear chassis section with rear-axle support,
19 Anti-roll bar with highly pretensioned anti-roll bar mounts,
20 17" sport brake.



Chassis design
A chassis in the classic sense is divided into the following subsystems:
– suspension (body suspension and anti-roll bars),
– shock absorbers,
– wheel suspension and suspension linkage (wheel control),
– wheels with tires,
– steering, and
– brake system.

Figure 1 shows a typical chassis and its components. The fundamentals and particulars of the individual subsystems of a chassis will be addressed in greater detail in the following.

Functions of the chassis
The fundamental function of the chassis is to move the car by transmitting the drive torque to the roadway. The driver influences the horizontal movement by steering; in the process, the vehicle should hold its course and allow the driver to guide the vehicle safely and comfortably.

The vertical movement of the vehicle is determined by the course of the road and road-surface irregularities, which excite vibrations. The chassis decisively determines these vibration effects occurring during driving in the longitudinal, lateral, and vertical vehicle directions. The function of the chassis is to reduce these vibrations in as controlled a manner as possible. It is essential to avoid as far as possible the transmission of these vibrations to the vehicle body so as to reduce rolling, yawing, and pitching and to prevent the bonanza effect (rocking).

Further functions of the chassis are to support the vehicle mass and to bear, guide, steer, and brake the wheels.

A dynamically well tuned chassis enables the driver to drive comfortably and safely because it implements the driver's inputs precisely. The driver can safely keep on top of or avoid entirely critical situations.

Lightweight construction
The chassis has a substantial influence on the weight. But the lightweight construction of the chassis components also plays a significant role in that the entire unsprung mass of the vehicle (wheels, tires, wheel carriers and wheel bearings, wheel suspension, springs, and shock absorbers) is located in the chassis. The lower unsprung mass of the chassis influences vehicle handling in a positive way; the lower it is, the less the wheel vibrations affect the body. This reduces disturbance variables and as a result increases driving safety and ride comfort.

Influence of driver-assistance systems
The driver-assistance systems assigned to active safety (e.g., driving-dynamics control, lane-keeping support) affect the stabilization and course holding of the vehicle via chassis components. This also illustrates the significance of the chassis.

Further chassis developments
Where the chassis in the past was characterized above all by mechanical components, so today active components (e.g., adjustable shock absorbers, rear-axle steering) and networked chassis systems are – above all in the passenger-car sector – state of the art. The interaction of active components, sensors for monitoring and calculating the driving condition, and intelligent control approaches open up a multitude of new possibilities for influencing relevant chassis properties in response to specific situations. These possibilities for the individual subsystems will also be addressed in greater detail in the following.

Basic principles

Dynamic behavior of a vehicle

Vertical movement

The roads normally used by motor vehicles feature irregularities that cover a frequency range of up to around 30 Hz and are the most intensive source of excitation in the vehicle. The resulting excitation leads to vertical movements (vertical accelerations) of the vehicle and its occupants. The interface between the road surface and the vehicle is the chassis system, the main task of which is power transmission between the environment and the vehicle body. This means that both the dynamic characteristics of the vehicle and driving safety (due to wheel-load fluctuations), as well as ride comfort (due to body acceleration), are heavily influenced by the choice of chassis system (Figure 2). Driving safety increases with decreasing effective wheel-load fluctuations, comfort increases with decreasing effective body accelerations.

Influencing the vehicle dynamics

The dynamics of the vehicle response, however, are not exclusively determined by the chassis system components, but rather are much more a consequence of the combination of different overall vehicle parameters. As a rule, influencing the vehicle dynamics with measures on the chassis system involves high complexity, in particular because the effects of parameter variations have to prevail among the conflicting priorities of driving safety and ride comfort.

Driving safety

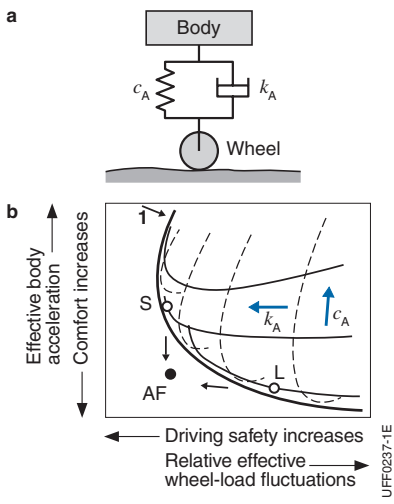
Driving safety depends decisively on the contact relationships between the tires and the road surface, and thus on the longitudinal and lateral forces that can be transferred. A fundamental aim in chassis configuration with regard to driving safety is therefore always to minimize the dynamic wheel load fluctuations that cause a reduction in the level of force that can be transferred.

Ride comfort

The ride comfort, on the other hand, depends on the movements and accelerations that affect the occupants (above all in the vertical direction). Depending on the area of application, comfort is of great relevance and should in no way be regarded merely as a concomitant of system development. In the case of professional drivers in particular, adequately high ride comfort is to be ensured in order to prevent long-term damage to health. The effective value of the body acceleration has proven to be a good evaluation parameter in this context.

Figure 2: Conflict diagram – area of conflict of driving safety and ride comfort

a) Suspension/shock-absorber system,
b) Conflict diagram.
1 Limit of conventional chassis.
k_A Spring constant,
c_A Dampening rate,
S Sports car, conventionally configured, passive,
L Sedan, conventionally configured, passive,
AF Active chassis, controlled.



Fundamental requirements for a chassis

These core parameters, however, initially only describe the potentials of a chassis system to meet the corresponding requirements. The actual dynamics of the vehicle, the driving safety, and the comfort also depend decisively on the choice of road-surface parameters (environment) and vehicle-internal manipulated variables (e.g., steering angle and accelerator-pedal position) set by the driver. The fundamental requirements for a chassis are shown in Figure 3 [1]. The chassis as a system (without electronics) is traditionally divided into the following subsystems:

– suspension,
– shock absorbers,
– wheel suspension and suspension linkage (wheel control),
– wheel-and-tire combination,
– steering, and
– brake system.

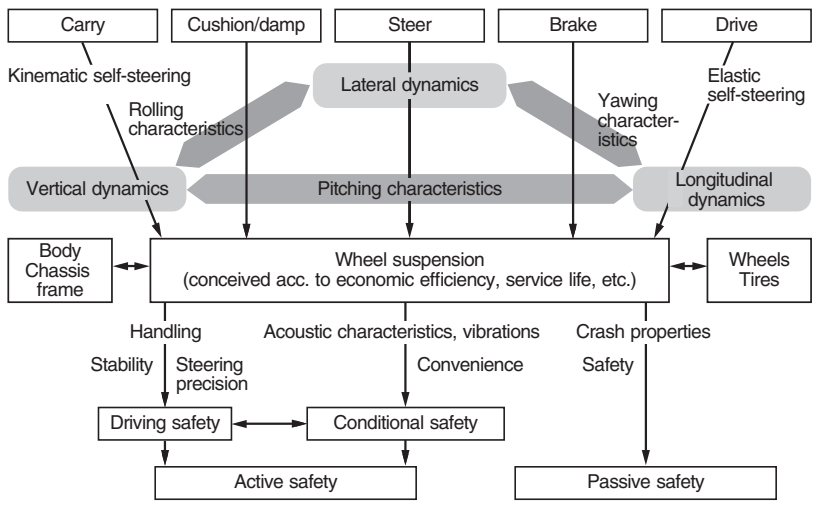
Alongside the adaptation of the oscillatory system to the external road excitations, vehicles also require the influence of internal sources of excitation on the vibration characteristics (drivetrain, wheel, and tire) to be minimized [2].

Knowledge and a description in the form of objective variables of the road excitation that causes vibrations are required in order to examine the vibration characteristics and to configure the suspension/shock absorber system of a chassis. (See Dynamics of vertical motion.) Whereas minor irregularities can already be compensated for by the suspension characteristics of the tire, an element between the wheels and body that changes its length is required to reduce greater body movements. Steel springs are used most frequently here, delivering a return force that depends on the change in length. The result is – taking account of the wheel and body masses – a system that can vibrate and requires other elements for damping.

Measures for vibration minimization

Road excitations lead to axle and body vibrations in vehicles. These are to be avoided in the course of minimizing dynamic wheel-load fluctuations (driving safety increases) and body accelerations (comfort increases). They are reduced

Figure 3: Requirements for a passenger-car chassis



by means of the corresponding components for suspension and damping within the wheel suspension, whereby different systems for suspension, damping, and suspension linkage have proven effective depending on the area of application. The influence of different suspension/shock absorber parameters and their effects on different frequency ranges is shown in Table 1.

Body springs and body dampers are traditionally used for these tasks. Coupled with kinematics and elastokinematics in the wheel suspension that are configured in line with requirements, these are intended to ensure optimal power transmission between the tires and road surfaces and, simultaneously, high levels of comfort. Sportier or comfortable tuning of the chassis can be achieved with passive elements, as the conflict diagram in Figure 2 shows. This conflict of aims can be resolved to a large extent by the use of active or adjustable-characteristic springs and dampers.

However, with the increasing use of actuators in the chassis area (e.g., superimposed steering, torque vectoring), the application overhead for optimization of the subsystems is becoming increasingly important, this being met among others by the integration of functions in a Central Chassis Control unit [3]. One of the most important reasons for this is the increase in the product added value due to low-cost function integration on the software side (i.e., the complete implementation of various functions of different subsystems in the software environment – without being forced to increase the technological overhead). In addition, the technical and economic synergy potentials in overall vehicle networking can be increased significantly and the differentiation potential of various derivatives can be enhanced using the same electronic and software modules. This means that electronics and software are becoming increasingly important in the area of chassis systems as complexity increases.

Table 1: Effects of changes to the suspension/shock-absorber system on the vertical vehicle vibration characteristics

	Design parameters	Effects in the low-frequency range (body natural frequency)	Effects in the middle frequency range	Effects in the higher-frequency range (axle natural frequency)
Body data	Spring constant	Body natural frequency and body acceleration drop radically when the spring rigidity is reduced	Body acceleration drops slightly on reduction of the spring rigidity	Body acceleration remains virtually constant on reduction of the spring rigidity
	Dampening rate	Body acceleration rises radically on reduction of the dampening rate	Body acceleration drops radically on reduction of the dampening rate	Dynamic wheel load fluctuation rises radically on reduction of the dampening rate
Tire data	Spring constant	Natural frequency and amplitude remain approximately constant		Natural frequency and amplitude of the body acceleration and wheel load fluctuations drop approximately proportionally to the reduction in vertical wheel spring rigidity
	Dampening rate	Natural frequency and amplitude remain approximately constant		Amplitude of the body acceleration and wheel load fluctuation drops slightly with increasing damping at consistent wheel natural frequency

Characteristics

Characteristics of the complete vehicle

The handling of the complete vehicle comprises essentially the behavior of the body – i.e., bodywork, passenger cell, engine, and transmission – and the suspension linkage (wheel control) – i.e., axle, steering, wheels, and tires. The motion of the body can generally be described as a rigid body. Even in the bodies of convertibles, deviations from rigid-body motion only occur above the frequency range of 10 to 15 Hz.

Characteristics of the complete vehicle, the wheel suspension, the steering kinematics, and the tire are defined for the purpose of describing the handling. The basic terms explained in the following (for quantities and units, see Table 2) describe vehicle handling in the important frequency range up to approx. 8 Hz. Many of these characteristics are described in the standards ISO 8855 [4] and DIN 70000 [5] (withdrawn standard).

Figure 4 shows the system of coordinates to which the layout of the components and quantities refer.

Wheel contact point

The wheel center plane and the road plane have an intersection line on which the wheel contact point is situated at the point where the connection of the wheel

Figure 4: System of coordinates
ψ Yaw angle, φ Roll angle, θ Pitch angle, S Center of gravity.

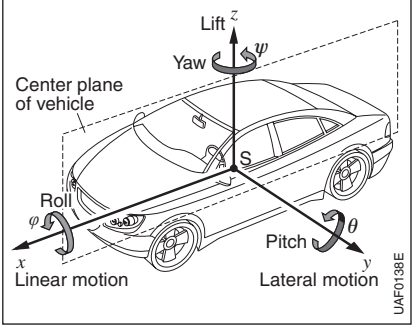


Table 2: Quantities and units

Quantity		Unit
α	Slip angle	degree
β	Float angle	degree
δ_H	Steering-wheel angle	degree
δ_R	Toe angle of right wheel	degree
δ_L	Toe angle of left wheel	degree
δ_A	Axle steering angle (steering angle)	degree
ε_V	Longitudinal inclination at wheel center	degree
ε_B	Anti-dive angle	degree
ε_A	Anti-lift angle	degree
ψ	Yaw angle	degree
φ	Roll angle	degree
θ	Pitch angle	degree
γ	Camber angle	degree
σ	Kingpin angle	degree
τ	Caster angle	degree
λ	Slip	–
λ_B	Braking-force distribution	–
ω_R	Angular velocity, wheel	s^{-1}
a_x	Longitudinal acceleration	m/s^2
a_y	Lateral acceleration	m/s^2
a_z	Vertical acceleration	m/s^2
a_t	Tangential acceleration	m/s^2
a_c	Centripetal acceleration	m/s^2
F_S	Lateral force	N
F_U	Longitudinal force	N
F_Z	Axle load	N
h	Height of center of gravity	m
h_v, h_h	Longitudinal-pole height, front/rear	m
h_W	Height of roll pole	m
i_s	Steering ratio	–
l	Wheelbase	m
M_H	Steering-wheel torque	Nm
M_R	Tire aligning torque	Nm
m	Vehicle mass	kg
n_τ	Caster offset	m
n_v	Caster offset in wheel center	m
n_R	Tire caster offset	m
r_σ	Kingpin offset at wheel center	m
r_{st}	Deflection-force lever arm	m
r_l	Scrub radius	m
r_{dyn}	Dynamic rolling radius	m
s	Track width	m
v_x	Longitudinal velocity	m/s
v_y	Lateral velocity	m/s
v_z	Vertical velocity	m/s
v_{RAP}	Speed at wheel contact point	m/s
v_{RMP}	Speed at wheel center	m/s
X_A	Starting-torque compensation	–
X_B	Braking-torque compensation	–

center to this intersection line intersects the intersection line vertically.

Wheelbase
The wheelbase l according to ISO 612 [6] is the distance between the wheel contact points of the front and rear wheels.

A long wheelbase delivers a lower pitch behavior and thereby facilitates better ride comfort. At the same time a long wheelbase reduces the yaw velocity which is generated by predefined wheel steering angle. A short wheelbase on the other hand delivers better maneuverability (e.g., when parking). The ratio of wheelbase to vehicle length is in the range of 0.6 to 0.7. A ratio of approximately 0.7 is established for small vehicles.

Track width
The track width s is the distance between the wheel contact points of the wheels on one axle.

A wide track width generally results in better handling and greater ride comfort. The drawbacks of a wider vehicle are the greater weight and the higher aerodynamic drag. Typical values for the ratio of track width to vehicle width are in the range of 0.80 to 0.86.

Center of gravity
The concentrated total mass of the vehicle is assumed at the center of gravity.

A low center of gravity results in good handling (typical of sports cars), low rolling and pitching.

Wheel load
The wheel load is the force acting on the wheel contact point between the roadway and the wheel vertically to the roadway.

When the vehicle is stationary on a level surface the sum total of the wheel loads is equal to the force due to weight of the complete vehicle.

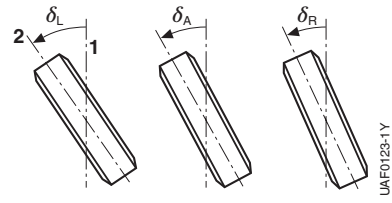
Axle-load distribution
The axle-load distribution specifies the ratio of the axle loads on the front and rear axles with regard to the total vehicle mass. For traction reasons on vehicles with front-wheel drive the axle load is 55 % on the front axle and accordingly 45 % on the rear axle. For rear-wheel drives the ratio can be established at 50:50.

Characteristics of the wheel suspension
The front wheels have two degrees of freedom, i.e., the degree of freedom for compression and rebound and that for steering. The rear wheels generally only have the one degree of freedom for compression and rebound. In the case of vehicles with four-wheel steering, the rear wheels also have the degree of freedom for steering. These degrees of freedom are defined by the kinematics and elastokinematics of the axle. Kinematics of the axle is the pure rigid-body motion of the individual suspension arms, while elastokinematics describes the behavior of the axle in response to the effect of forces and moments. The characteristics of the wheel suspension are described in the following.

Toe angle
Moving the steering wheel causes the two front wheels to turn in the same direction. This alters the toe angle of the right and left wheels (δ_R and δ_L) respectively. The toe angle is the angle between the vehicle center plane and the wheel center plane when projected onto the x-y plane (roadway, Figure 5). In the case of a positive turn, i.e., in the counterclockwise direction about the wheel's z-axis, the toe angle is also positive.

The difference between the toe angles can amount to several degrees with the steering-wheel angle at maximum. When viewed geometrically, the toe angle of the

Figure 5: Toe angle and axle steering angle
1 Vehicle center plane,
2 Wheel center plane.
δ_L Toe angle of left wheel,
δ_R Toe angle of right wheel,
δ_A Axle steering angle (steering angle), mean toe angle.



inside wheel is greater than the toe angle of the outside wheel.

On vehicles with four-wheel steering the rear wheels also move.

Axle steering angle (steering angle)
The mean toe angle is called the axle steering angle δ_A or simply the steering angle (Figure 5).

Toe-in and toe-out
When the steering wheel is in the straight-ahead position, the toe angles of the front wheels are in the dynamically advantageous range of 0.1 to 0.3°. When the distance of the rim flanges ahead of the wheel centers is smaller than that of the rim flanges after the wheel centers, this is referred to as wheel toe-in (Figure 6). When the situation is reversed, this is referred to as wheel toe-out. Both toe-in and toe-out are given in degrees (°).

The terms toe-in and toe-out are also used to refer to a single wheel. In this case, toe-in means that the wheel has a toe angle in the direction of the vehicle center plane and toe-out that it has a toe angle against the direction of the vehicle center plane.

Camber angle
The camber angle γ is the angle between the vehicle center plane and the wheel center plane when projected onto the z-y plane. The camber angle is positive when the wheels are further away from the vehicle center plane at the top than at the bottom (Figure 7a). On account of the axle kinematics the camber angle relative to the body is dependent on the spring travel. The spring travel is the travel obtained when the wheel rebounds and compresses relative to the body referred to the z-direction.

Further to this definition, the camber angle relative to the road is likewise important to vehicle handling. The camber angle relative to the road is the angle between the wheel center plane and the normal line to the road surface (Figure 7b). The preceding sign is determined according to the rectangular right-handed system. If the vehicle center plane is perpendicular to the road surface, both definitions of the camber angle are equal in terms of amount. Otherwise, it is important to observe the precise definition.

Transverse pole
When the axle is subjected to compression and rebound the position of the wheels is determined primarily by the kinematics and elastokinematics. The wheel moves transversally to the direction of travel about the transverse pole (Figure 8). The speeds for

Figure 6: Toe-in
1 Wheel center.
d_v Distance between rim flanges, front,
d_h Distance between rim flanges, rear,
δ_L Toe angle of left wheel,
δ_R Toe angle of right wheel,
⇨ Direction of travel.

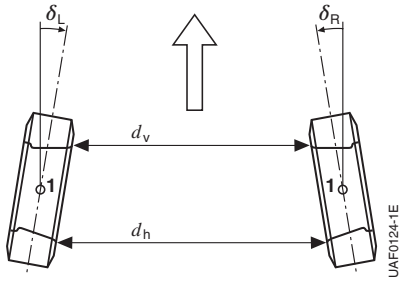
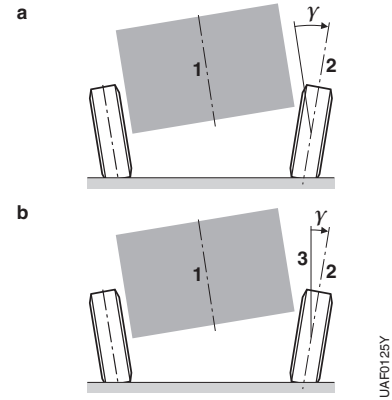


Figure 7: Camber
a) Relative to the body,
b) Relative to the road.
1 Vehicle center plane,
2 Wheel center plane,
3 Normal line to the road surface.
γ Camber angle.



example at the wheel contact point (v_{RAP}) and at the wheel center (v_{RMP}) are during compression and rebound perpendicular on the connecting line to the transverse pole. The position of the transverse pole changes during compression and rebound.

Roll pole
The vehicle body moves in the case of low lateral accelerations about the roll pole of the respective axle (Figure 8). The roll pole is on the connecting line for wheel contact point and transverse pole in the vehicle center plane, i.e., at half the track width (s/2). The height h_W of the roll pole can thus be easily calculated:

$$h_w = \frac{v_{RAP,y}}{v_{RAP,z}} \frac{s}{2} .$$

The height of the roll pole is typically below 120 mm. In order in the case of high lateral accelerations to avoid the support effect – i.e., the jacking effect – the height of the roll pole decreases with compression.

The roll pole is also known as the roll center or instantaneous center.

Roll axis
The connection between the roll pole of the front axle and the roll pole of the rear axle is called the roll axis (Figure 9). The center of gravity of the body is usually above the roll axis. Center-of-gravity heights for sedans range between 550 and 650 mm. This roll axis applies for lower lateral accelerations. In the case of higher lateral accelerations, both the suspension adjustment and the axle behavior must be taken into consideration. The roll axis is then not inevitably in the vehicle center plane.

Longitudinal inclination at wheel center
The angle between the direction of motion of the wheel center during compression and rebound and the z-axis is called the longitudinal inclination at wheel center ε_V (Figure 10).

Longitudinal pole
When the spring movement of the axle is projected onto the vehicle center plane, the wheel moves about the longitudinal pole L (Figure 10), i.e., the wheel center moves during compression not along the z-axis but along the direction which was rotated about the longitudinal inclination at wheel center relative to the z-axis. The speeds at the wheel contact point (v_{RAP}) and the wheel center (v_{RMP}) are during compression and rebound perpendicular on the connecting line to the longitudinal pole. The position of the longitudinal pole can change during compression and rebound. The angle between the connection of the wheel contact

Figure 8: Transverse pole and roll pole
1 Wheel center, 2 Tire center plane,
3 Vehicle center plane.
Q Transverse pole, W Roll pole.
s Track width, h_w Height of roll pole.
v_{RAP} Speed at wheel contact point,
v_{RMP} Speed at wheel center.

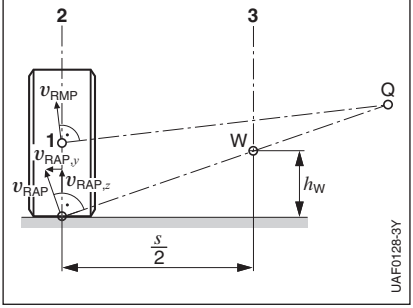
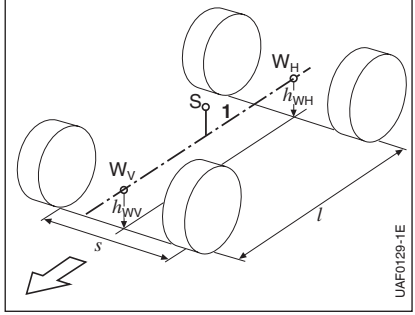


Figure 9: Roll axis
1 Roll axis.
S Center of gravity, W_V Roll pole, front axle,
W_H Roll pole, rear axle,
s Track width, l Wheelbase,
h_{WV} Roll-pole height, front axle,
h_{WH} Roll-pole height, rear axle.
⇦ Direction of travel.

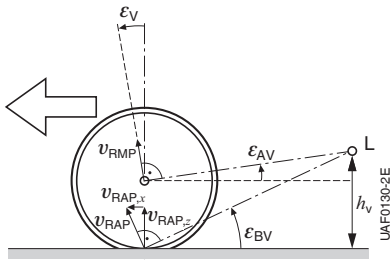


point to the longitudinal pole and the road is called the anti-dive angle ε_B. The angle between the connection of the wheel center to the longitudinal pole and the parallel line to the road is called the anti-lift angle ε_A.

The longitudinal pole of the front axle is behind the front wheels and the longitudinal pole of the rear axle is in front of the rear wheels (Figure 11).

Figure 10: Longitudinal pole, anti-lift angle and anti-dive angle

L Longitudinal pole,
h_V Longitudinal-pole height, front,
ε_V Longitudinal inclination at wheel center,
ε_{BV} Anti-dive angle, front,
ε_{AV} Anti-lift angle, front,
v_{RAP} Speed at wheel contact point,
v_{RMP} Speed at wheel center.
◁ Direction of travel.



When the vehicle is braked, the front axle load is increased by ΔF_z and the rear axle load decreased by ΔF_z. At the vehicle center of gravity the force $F = m\,a$ is applied. With braking-force distribution λ_B the braking forces $F_{xV} = \lambda_B F_x$ and $F_{xH} = (1 - \lambda_B) F_x$ are applied to the front and rear axles respectively. In the optimum case where the resulting force from F_{xV} and ΔF_z on the front axle passes exactly through the longitudinal pole and similarly on the rear axle, no spring movement occurs in the vehicle's body springs. For the optimum anti-dive and anti-rise angles:

$$\tan(\varepsilon_{BV,opt}) = \frac{1}{\lambda_B}\,\frac{h}{l},$$

$$\tan(\varepsilon_{BH,opt}) = \frac{1}{(1-\lambda_B)}\,\frac{h}{l},$$

with wheelbase l and height of center of gravity h.

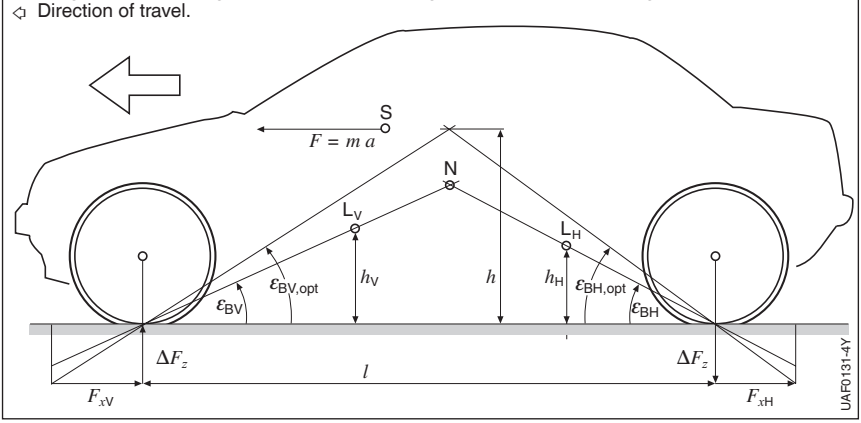
Similarly, for the case of traction, an optimum anti-lift angle can be calculated in each case on the front and rear axles. In this case λ_A specifies the proportion of tractive force on the front axle. For the optimum anti-lift angle on the front and rear axles:

$$\tan(\varepsilon_{AV,opt}) = \frac{1}{\lambda_A}\,\frac{h}{l},$$

$$\tan(\varepsilon_{AH,opt}) = \frac{1}{(1-\lambda_A)}\,\frac{h}{l}.$$

Figure 11: Longitudinal pole and pitch pole

S Center of gravity, N Pitch pole, L_V Longitudinal pole, front axle, L_H Longitudinal pole, rear axle, m Vehicle mass, a Acceleration, F_{xV} Braking force on front axle, F_{xH} Braking force on rear axle, ΔF_z Axle-load change, l Wheelbase, h Height of center of gravity, h_V Longitudinal-pole height, front, h_H Longitudinal-pole height, rear, ε_{BV} Anti-dive angle, front, ε_{BH} Anti-dive angle, rear.
◁ Direction of travel.



Pitch pole
The two connecting lines between the wheel contact points and the longitudinal pole intersect at the pitch pole N (Figure 11). The pitch pole N is always below the height of the center of gravity so that an expected vehicle pitching motion towards the front can always be effected when braking.

Pitch axis
The pitch axis passes through the pitch pole and is perpendicular to the vehicle center plane.

Braking-torque compensation and starting-torque compensation
Braking-torque compensation X_{BV} and X_{BH} is a measure of how effectively optimum braking compensation has been implemented. The following applies:

$$X_{BV} = \frac{\tan(\varepsilon_{BV})}{\tan(\varepsilon_{BV,opt})}\ 100\ \%\ ,$$

$$X_{BH} = \frac{\tan(\varepsilon_{BH})}{\tan(\varepsilon_{BH,opt})}\ 100\ \%\ .$$

The same applies to acceleration. It is necessary to take into account firstly the type of drive, i.e., four-, front- or rear-wheel drive, and secondly the fact that the tractive force is applied at the wheel center. In the general case of four-wheel drive the following applies to starting-torque compensation X_{AV} (front) and X_{AH} (rear):

$$X_{AV} = \frac{\tan(\varepsilon_{AV})}{\tan(\varepsilon_{AV,opt})}\ 100\ \%\ ,$$

$$X_{AH} = \frac{\tan(\varepsilon_{AH})}{\tan(\varepsilon_{AH,opt})}\ 100\ \%\ .$$

Characteristics of steering kinematics

Steering-wheel angle and steering-wheel torque
The steering-wheel angle δ_H is the turning angle of the steering wheel measured from the straight-ahead position. The angle is positive for a left-hand curve.

The driver must apply a torque when adjusting the steering-wheel angle. This torque is called the steering-wheel torque M_H and is also positive for a left-hand curve.

Steering ratio
Essentially, on account of the steering gear but also on account of the kinematics of the front axle, the toe angle obtained at the wheels is significantly smaller than the steering-wheel angle. In such a case where there are no acting forces or torques and where the vehicle is barely loaded, the steering ratio i_s is governed by the following equation:

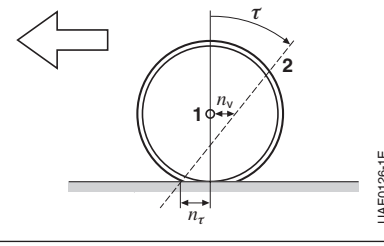
$$i_s = \frac{2\ \delta_H}{(\delta_L + \delta_R)}\ .$$

Steering axis (kingpin axis)
During steering the wheels move not about their z-axis, but rather about the steering axis, also known as the kingpin axis. The position of the kingpin axis is essentially determined by the kinematics of the axle (Figures 12 and 13).

Steering-axis spread (kingpin angle)
When the wheel and the kingpin axis are projected onto the vehicle transverse plane (y-z plane), the kingpin axis is inclined by the kingpin angle σ (Figure 13). The kingpin angle (also known as the kingpin inclination) is positive when the kingpin axis is inclined to the vehicle center. Kingpin angles are generally positive.

Kingpin offset at wheel center
The distance between the wheel center and the kingpin axis parallel to the road is called the kingpin offset at wheel center r_σ. The kingpin offset at wheel center is positive when the wheel center is further from the vehicle center plane than the kingpin axis (Figure 13).

Figure 12: Position of kingpin axis when projected onto the vehicle center plane
1 Wheel center, 2 Kingpin axis.
τ Caster angle,
n_v Caster offset in wheel center,
n_τ Caster offset.
◁ Direction of travel.



Deflection-force lever arm
The shortest connection between the wheel center and the kingpin axis is called the deflection-force lever arm r_{st}. The deflection-force lever arm is positive when the wheel center is further from the vehicle center plane than the kingpin axis.

The axes are arranged in such a way that the kingpin offset at wheel center and the deflection-force lever arm are as small as possible. This arrangement prevents deflection forces from being generated in the steering.

Scrub radius
The distance between the wheel contact point and the point where the kingpin axis intersects the road surface is called the scrub radius r_l (Figure 13). The scrub radius is positive when the wheel contact point is further from the center plane than the kingpin axis. When the scrub radius is positive, the wheel moves towards toe-out under braking forces. This behavior is particularly advantageous during the "braking in a curve" maneuver. When the scrub radius is negative, the wheel moves towards toe-in under braking forces. During the "μ-split braking" maneuver with different coefficients of friction on the left and right sides of the vehicle this axle behavior creates the precondition for more stable vehicle handling. On account of these different effects the scrub radius is configured to be as small as possible.

Caster angle, caster offset in wheel center and caster offset
When the wheel and the kingpin axis are projected onto the vehicle center plane, the kingpin axis is inclined by the caster angle τ. It is positive when the top end of the kingpin axis is inclined towards the rear (Figure 12). In this projection the kingpin axis generally does not pass through the wheel center; instead, it is offset by the caster offset in wheel center n_v to the rear.

The distance between the wheel contact point and the point where the kingpin axis intersects the road surface is called the caster offset n_τ. If this point of intersection is located in front of the wheel center (Figure 12), the caster offset is positive. Caster offsets are typically in the range of 15 to 30 mm.

Rack travel
The most widely used type of steering today is rack-and-pinion steering. Here the steering-wheel angle is converted via a pinion into a translational motion of the rack (in the direction of the y-axis). The rack is connected at both ends to the front-axle tie rods and thereby enables the toe angle to be adjusted on both wheels. The

Figure 13: Position of kingpin axis when projected onto the vehicle transverse plane
1 Wheel center, 2 Kingpin axis,
3 Vehicle center plane, 4 Wheel contact point.
σ Kingpin angle,
r_σ Kingpin offset at wheel center,
r_l Scrub radius, r_{st} Deflection-force lever arm.

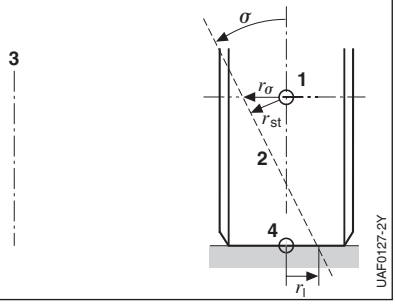
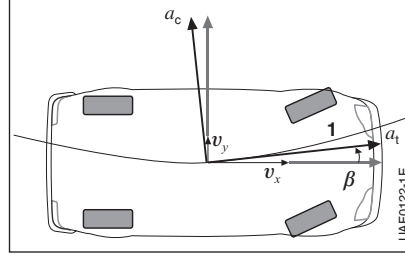


Figure 14: Float angle, centripetal and tangential acceleration
1 Trajectory.
β Float angle,
a_t Tangential acceleration,
a_c Centripetal acceleration,
v_x Longitudinal velocity,
v_y Lateral velocity.



distance covered by the rack when steering is called the rack travel.

Float angle
In the dynamics of lateral motion the center of gravity of the body does not always move along the x-axis. The angle that forms between the vehicle center plane and the trajectory is called the float angle β (Figure 14). It is counted from the vehicle center plane to the trajectory. The float angle is calculated from the linear velocity v_x and the lateral velocity v_y:

$$\beta = \arctan \frac{v_y}{v_x}.$$

Because linear velocity during forward driving is by definition positive, the sign preceding the lateral velocity in this case also determines the sign preceding the float angle.

Ackermann angle
The Ackermann angle corresponds to the axle steering angle (Figure 5). In the event of very low lateral acceleration a_y, i.e., if virtually no lateral force is built up, the vehicle moves along the Ackermann angle.

Turning circle
The turning circle is described when cornering at full steering lock by the farthest protruding vehicle parts. The radius of the circle formed is the turning circle.

The turning circle is dependent on the wheelbase and the maximum steering lock, i.e., on the maximum toe angle on the front axle.

Basic terms for tires
The most important external forces and moments which act on a vehicle occur during the transmission of force between tire and road surface. Added to this are the wind forces which act on the vehicle special situations.

Footprint
Force is transmitted between tire and road surface by friction in the contact area, also known as the footprint. The two most important types of friction are adhesive friction (intermolecular adhesive force) and hysteretic friction (gearing force).

Lateral force and longitudinal force.
Forces can be generated in the road plane by friction. Lateral force F_S is the force perpendicular to the wheel center plane; longitudinal force F_U runs in the direction of the wheel center plane (Figure 15). The forces are generally not applied exactly at the wheel contact point such that moments are likewise generated in relation to the wheel contact point (tire aligning torque M_R).

Slip angle
When lateral and longitudinal forces occur simultaneously, the forces may be mutually influenced. The situation where lateral and longitudinal forces occur but not combined is considered in the following. The gener-

Figure 15: Lateral force and longitudinal force. (View from above).
1 Wheel contact point.
F_U Longitudinal force,
F_S Lateral force,
M_R Tire aligning torque.

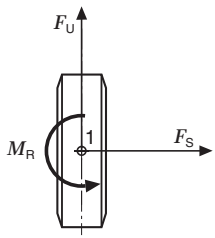
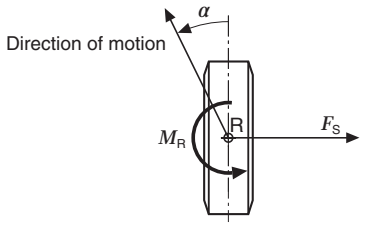


Figure 16: Slip angle
(View from above).
F_S Lateral force,
M_R Tire aligning torque,
α Slip angle.
R Wheel contact point.



ated lateral force F_S is dependent on the wheel load and the slip angle α. A dependence of lateral force F_S on speed can generally be ignored. The wheel load is the force with which the wheel center is pressed towards the road-surface plane. The slip angle α is the angle between the direction of motion of the wheel contact point and the wheel center plane (Figure 16).

When the wheel load is kept constant and the slip angle α is increased, the lateral force initially increases linearly. The lateral force reaches its maximum level at a slip angle of approx. 5° and then decreases slightly (Figure 17).

Tire aligning torque and tire caster offset
With small slip angles the lateral force is applied after the wheel contact point. When the slip angle is increased, the lateral force moves increasingly towards the wheel contact point and can also be located ahead of the wheel contact point. The distance between the lateral application point and the wheel contact point is called the tire caster offset n_R. The lateral force therefore generates a moment about the tire's vertical axis, the tire aligning torque M_R:

$$M_R = F_S \, n_R \, .$$

This produces, at constant wheel load, a curve as shown in Figure 18 for the tire aligning torque.

If the tire aligning torque is positive, it helps to make the slip angle smaller in terms of amount. This behavior helps to make the turned wheels return to the straight-ahead position when the steering wheel is released.

Slip and rolling radius
Similarly to the slip angle α for lateral force F_S, slip λ is the variable which at constant wheel load determines longitudinal force F_U. Slip occurs when the speed v_{xR} at which the wheel center moves in the longitudinal direction differs from the speed v_U at which the circumference rolls. The circumferential speed is calculated from the angular speed ω_R of the wheel and the dynamic rolling radius r_{dyn}:

$$v_U = \omega_R \, r_{dyn} \, .$$

A distinction is made between static and dynamic rolling radius. The static rolling radius is the shortest distance between wheel center and footprint. The dynamic rolling radius r_{dyn} is calculated by means of the circumference U:

$$r_{dyn} = \frac{U}{2\pi} \, .$$

Slip λ_A for tractive forces is defined as:

$$\lambda_A = \frac{\omega_R \, r_{dyn} - v_{xR}}{\omega_R \, r_{dyn}} \, .$$

Similarly, slip λ_B for braking forces is defined as:

$$\lambda_B = \frac{\omega_R \, r_{dyn} - v_{xR}}{v_{xR}} \, .$$

Figure 17: Lateral force at constant wheel load as a function of slip angle

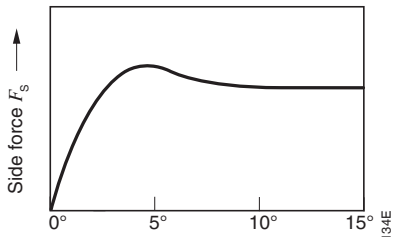
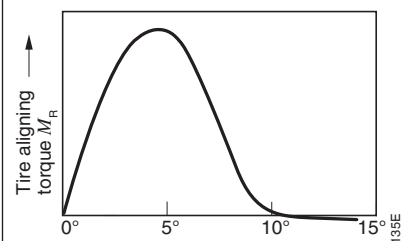


Figure 18: Tire aligning torque at constant wheel load as a function of slip angle



According to this definition, drive slip is always positive and brake slip always negative. These two slip definitions ensure that when a wheel is locked ($\omega_R = 0$, $\lambda_B = -1$) a slip of –100 % is obtained and when a wheel is spinning ($v_{xR} = 0$, $\lambda_A = 1$) a slip of 100 % is obtained.

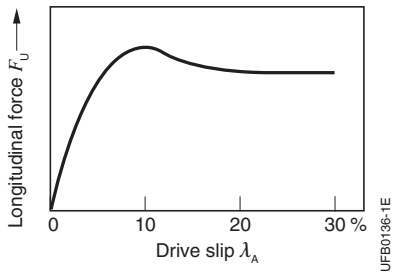
If the drive slip is increased at constant wheel load, the tractive force (longitudinal force) increases linearly. At approx. 10 % drive slip the longitudinal force reaches its maximum level and then falls again (Figure 19). The same applies to brake slip. Here the maximum braking force is created at approx. –10 %.

Measured variables

In the interests of vehicle development it is sensible to use a special measurement to enable vehicle handling to be measured with greater accuracy. In the dynamics of lateral motion the translational accelerations and the attitude angles are often measured by gyro-stabilized platforms. The absolute positions are recorded with GPS measuring systems. In the dynamics of vertical motion the translational accelerations are measured in each of the three directions in space at different points of the body. From these the most important body accelerations can be determined, i.e., those for lift, pitch and roll.

The linear and lateral velocities are each measured with proximity-type velocity sensors for the float angle.

Figure 19: Longitudinal force at constant wheel load as a function of drive slip



The slip angle α is measured, along similar lines to the float angle, with two proximity-type speed sensors. The wheel speed and the longitudinal velocity are measured for slip λ. The dynamic rolling radius r_{dyn} is determined on test benches.

The tire aligning force M_R, the lateral force F_S, and the longitudinal force F_U can be recorded in mobile operation by multi-component measuring wheels. Because this is very costly, tire forces and tire moments are measured on stationary test benches or determined using special vehicles directly on the road. Up to now, precise measurement of tire forces and tire moments has been accompanied by many systematic failures.

Steering-wheel angle and steering-wheel torque are measured using special measuring steering wheels. If the measuring accuracy is sufficient, the steering-wheel-angle sensor fitted as standard in many vehicles can also be used.

Special measuring devices can be used in both mobile and stationary applications to measure toe and camber angles. These angles are typically measured on special test benches.

Special test benches are also used to measure the axle steering angle and the steering ratio.

The variables for the position of the kingpin axis are generally not measured directly. The axle pivot points are often recorded by means of geometric measurements and the following variables are calculated from them: caster angle, caster offset in wheel center, caster offset, kingpin angle, kingpin offset at wheel center, deflection-force lever arm, and scrub radius.

The roll pole can be determined by measuring the track change during the reciprocal compression and rebound of an axle. The roll axis is obtained from the roll poles of the front and rear axles.

The pitch pole and starting- and braking-torque compensation are usually not measured directly, but rather determined from the measured kinematic points of the axle. The same applies to the longitudinal inclination at wheel center, and the anti-lift/anti-squat and anti-dive/anti-rise angles.

References
[1] M. Ersoy, S. Gies (Editors): Fahrwerkhandbuch. 5th Ed., Verlag Springer Vieweg, 2017.
[2] M. Mitschke, H. Wallentowitz: Dynamik der Kraftfahrzeuge. 5th Edition, Verlag Springer Vieweg, 2015.
[3] J. Jablonowski, et al.: The Chassis of the all-new Audi A8. Chassis.tech 2017, Munich (2017).
[4] ISO 8855: Road vehicles – Vehicle dynamics and road-holding ability – Vocabulary.
[5] DIN 70000 (earlier standard) Road vehicles; vehicle dynamics and road-holding ability; vocabulary.
[6] ISO 612: Road vehicles; Dimensions of motor vehicles and towed vehicles; Terms and definitions.

Suspension

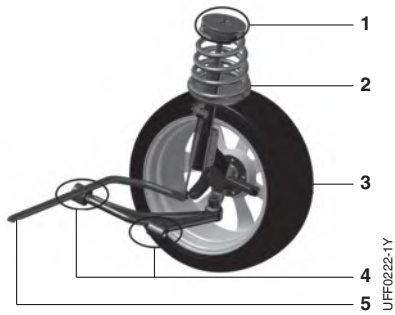
Basic principles

The suspension system of a vehicle has a decisive influence on the vibration characteristics and therefore on both comfort and driving safety. Depending on the vehicle category and use case, different solutions have prevailed in the meantime. An overview of the different suspension design elements is shown in Figure 1, using a quarter vehicle as an example.

As a general principle, suspension design elements include all parts of the wheel suspension of a motor vehicle that deliver return forces in the case of elastic deformation. The media that perform the suspension work on the different suspension systems are either steel (spring steel), polymer materials (rubber), or a gas (air).

Tires
As the connecting element between the road surface and vehicle, the tire is the first suspension design element in the transfer chain from excitation to the occupants that has a decisive influence on both the comfort (acoustics, rolling characteristics) and driving safety (longitudinal and lateral force potential). It has both suspension and damping properties, whereby these are not sufficient to eliminate the need for other vibration-absorbing elements in modern vehicles. An exception here is mobile machinery, where changed vertical-dynamics requirements mean that in almost all cases the suspension and damping are achieved by means of the tires.

Figure 1: Suspension design elements within the wheel suspension (McPherson split axle as an example)
1 Dome mount (rubber mount),
2 Body spring, 3 Tire,
4 Rubber mounts, 5 Stabilizer.



Elastomer mounts
Elastomer mounts are rubber elements with different functions and properties that interconnect individual components of a chassis system or secure them to the body.

The rubber mounts are used to provide insulation against vibrations and thus enhance comfort, particularly in the case of higher-frequency excitation (acoustics). At the same time the driving dynamics can be decisively influenced by the elasto-kinematics.

In contrast to series-production vehicles, uniball joints (more rigid connections between wheel suspension and body) are deployed in motor racing; this improves driving dynamics to the detriment of comfort.

In order to reconcile the conflicting objectives of soft mounts for high comfort and rigid mounts for sporty driving dynamics, in the meantime increasing use is made of adaptive or active suspension mounts that are able to adapt their properties to each driving situation.

Body springs
Body springs are parts of the chassis system that provide most of the vertical return forces between the wheel and body. Depending on the use case, various types of spring with very varied properties are used. A summary of the characteristics of suspension design elements used in vehicle construction can be found in Table 1.

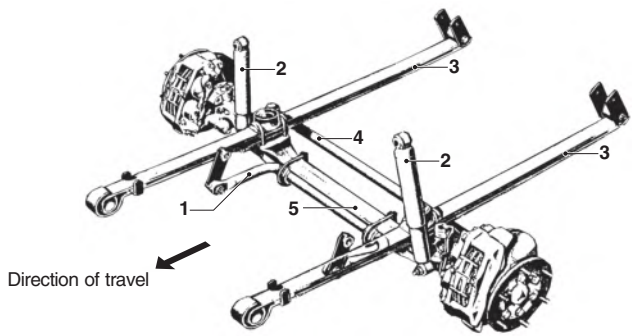
Table 1: Suspension design elements in vehicle construction

Suspension design element	Load influence on body natural frequency	Properties (●), advantages (+) and disadvantages (–)
Steel springs		
Leaf spring	● Natural frequency drops with increasing load ● Characteristic curves are generally linear or progressive	● Assumption of the suspension-linkage function is possible ● Single-layer or multilayer version ● Depending on type, subject to friction (in the passenger car an intermediate plastic layer reduces friction → positive influence on acoustics) + Good transfer of forces to the chassis (on trucks) + Low cost – Maintenance requirement – Friction damping usually inadequate – Acoustic influences
Coil spring		● Arrangement of the shock absorbers within the spring is possible ● Progressive characteristic curve due to corresponding geometry of the spring can be implemented (variable pitch or conical wire) + Great leeway for configuration + Low cost + No intrinsic damping + Low space requirement + Low weight + Maintenance-free – Separate elements required for suspension linkage – Spring characteristic curve is non-variable
Torsion-bar spring		● Made of round steel (for lower weight) or flat steel (with increased stress) + Wear- and maintenance-free + Depending on design, ride-height adjustment also possible – Long springs – Wheel-related spring stiffness depends on suspension-arm arrangement
Stabilizer	● No influence with suspension on the same side ● Half of stabilizer rigidity effective with one-sided suspension ● Entire stabilizer rigidity effective with alternating suspension	● Influence on self-steering effect (oversteer or understeer) ● U-shaped, bent fully round or tube material is usual ● Stub often flat-rolled due to bending stress ● Stabilizer attachment points located far out on the axle to achieve small diameters ● Axes of rotation of suspension arms configured so that stabilizer load is only torsion (not flexion) + Simple possibility to influence driving dynamics of a vehicle + Reduction of the roll angle + With deployment of active systems improved comfort and enhanced driving dynamics – Additional weight – Costs

Table 1: Suspension design elements in vehicle construction (continued)

Suspension design element	Load influence on body natural frequency	Properties (●), advantages (+) and disadvantages (–)
Air springs and hydropneumatic springs		
U-type bellows gas springs and air springs with bellows (air springs)	● Natural frequency remains constant with increasing load ● Characteristic curves depend on gas properties, form of rolling piston shape, and cord angle in the bellows	● Implementation of soft, vertical spring stiffness ● As spring strut or individual spring, can be found above all in commercial vehicles and buses ● Increasingly used in passenger cars for ride-height control of the rear axle + Ride-height control easy to integrate (similar to hydropneumatic spring) + Comfort characteristics independent of payload – Separate elements required for suspension linkage – Low pressure (< 10 bar) requires high volumes
Hydropneumatic spring	● Natural frequency rises with increasing load due to nonlinear spring stiffness	● Gas volume in the spring accumulator determines the suspension properties ● Power flux by means of gas and oil ● Integration of the damper valves in the shock absorber and in the connection between the suspension strut and the accumulator + Hydraulic damping and ride-height control easy to integrate – Maintenance requirement for rubber membrane due to diffusion tendency
Rubber springs		
Rubber spring	● Natural frequency is influenced by increased load due to nonlinear spring stiffness	● Vulcanized rubber shear spring between metal parts ● Used and assembly mountings (engine and transmission), suspension-arm mountings, additional springs, etc. ● Increasingly with integrated hydraulic damping + Design is very freely definable + Low cost – Limited temperature range – Aging

Figure 2: Example of a leaf spring with suspension-linkage function
1 Stabilizer, 2 Damper, 3 Leaf springs, 4 Panhard rod, 5 Rigid axle.



Types of spring

Leaf springs

The oldest types of spring used in vehicle construction are leaf springs, which were even used on horse-drawn coaches (Figure 2). Alongside the suspension functionality, a major advantage of this type of spring is its possible use as a suspension-linkage engineering-design element to connect the body and axle, because, with the appropriate design, it can absorb lateral or longitudinal forces. Multilayer leaf springs also have damping characteristics which, however, can lead to poor response and acoustic influences. The damping forces that can be achieved are not sufficient to completely eliminate the need for conventional shock absorbers.

The influences on comfort and their weight mean that leaf springs in the meantime no longer meet market requirements for personal transportation and they are therefore only used now in a few passenger cars (light utility vans, off-road vehicles). It is still usual to use this type of spring in the commercial-vehicle sector on account of the low costs and high reliability.

Helical springs

The great leeway for configuration with simultaneously low costs means that helical springs are the most frequently deployed type of body springs in the field of passenger cars. On this type of spring, the return forces are generated by the elastic torsion of individual coils during the change in length.

As helical springs are mainly able to absorb forces in the direction of the longitudinal axis of the spring, when they are used as body springs the other force components must be braced by the suspension linkage.

A geometric configuration of the springs that is in line with requirements (wire thickness, coil diameter and spacing, Figure 3) means that not only different design envelopes but also different spring characteristics can be achieved. This can be used in turn to influence the load-dependent body natural frequency and thus the ride comfort.

Torsion-bar springs

This type of spring is encountered mainly in passenger cars and light utility vans. They are bars made of spring steel that are subjected to torsion. The fixed clamping of one end of the bar and the rotatable mounting of the other end means that the shaft is elastically torsioned when subjected to a load in the form of torque applied in the direction of its axis. In the motor vehicle, the elastic twisting of the torsion bar is achieved with the help of a crank secured to the rotatable end of the rod (Figure 4). As a rule, the crank arms are suspension arms of the axle or wheel suspension. The torsion-bar springs are usually arranged in the bearing center of the suspension arms on the body side on the opposite end of which the vertical wheel force F_R takes effect as the external load.

Figure 3: Design examples of different types of helical springs
a) Changeable coil diameter,
b) Changeable wire diameter,
c) Changeable coil spacing,
d) Mini-block spring (combination of a, b, and c).

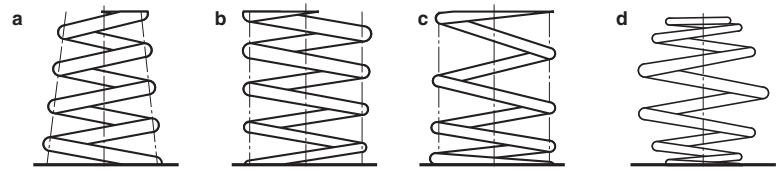


Figure 4: Design of a torsion-bar spring
1 Connection with the body,
2 Wheel connection.
l Length of torsion bar,
r Length of crank arm,
F_R Wheel force,
z_R Compression path,
ψ Torsion angle.

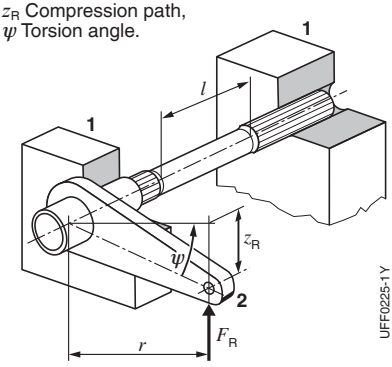
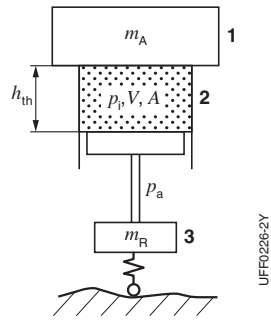


Figure 5: Structure of a gas spring
1 Structure, 2 Gas spring, 3 Wheel.
h_{th} Theoretical spring length,
m_A Body mass,
m_R Wheel mass,
p_i Pressure in the gas spring (internal pressure),
p_a Ambient pressure,
V Working volume of the gas spring,
A Area to which gas pressure is applied.



Gas springs

The body springs presented so far are fixed, springing media, whereby the work is performed by the change in shape of the steel springs. In contrast to this, the spring work in the case of gas springs is generated by a change in volume of the gas. The body of the vehicle is decoupled by an effective gas volume (possibly also by an additional fluid, see hydropneumatic springs) from the excitation and it vibrates on the gas cushion within the gas spring (Figure 5). This results in a favorable possibility to integrate a ride-height control function that can be implemented by pumping the intermediate medium (gas or fluid) in or out.

A characteristic parameter of the gas spring is the "theoretical spring length" h_{th}, which results as a quotient from the compression-dependent working volume $V(z)$ (including any additional volume) and the effective piston surface area A to which gas pressure is applied:

$$h_{th} = \frac{V(z)}{A} \qquad \text{(Eq. 1)}$$

where
z Compression path.

The equation for the spring force F:

$$F = (p_i - p_a)\, A \qquad \text{(Eq. 2)}$$

where
p_a Ambient pressure, and
p_i Internal pressure,

generally leads to the following for the spring stiffness of a gas spring with constant effective piston surface area A [1]:

$$c(z) = A\, n\, p(z)\, \frac{1}{h_{th}} \qquad \text{(Eq. 3).}$$

The polytropic exponent for slow and thus isothermic spring movements is $n = 1$ – for fast and thus approximately adiabatic spring movements it is $n = 1.4$. The natural angular frequencies for single-mass oscillators are calculated from

$$\omega_{Gas} = \sqrt{\frac{c}{m}} = \sqrt{\frac{c(z)\, g}{(p - p_a) A}}$$

$$= \sqrt{\frac{g\, n\, p(z)}{(p - p_a)\, h_{th}}} \qquad \text{(Eq. 4).}$$

When the requirement for a relatively small spring diameter is met $p_i >> p_a$. This simplifies the equation for the natural angular frequency to:

$$\omega_{Gas} = \sqrt{\frac{g\,n}{h_{th}}} \qquad \text{(Eq.5)}.$$

However, the above-mentioned theoretical piston cylinder gas spring is only used in vehicles in a modified form, whereby in principle a distinction is made between two types of gas spring, the air spring with bellows and the hydropneumatic spring. The fundamental difference with regard to the vertical dynamics lies in the influence of the load on driving smoothness and in the different effects on the spring stiffness in the level balancing of both systems. Whereas in the case of the hydropneumatic type level balancing is achieved by pumping in fluid or oil (with constant mass of the gas in the spring), level balancing on air springs with bellows takes place by pumping a gas (air) into the spring, thus restoring the original suspension volume. The change in spring stiffness of the hydropneumatic spring type that this causes leads to an increase in the body natural frequency with increasing load. In contrast to this, the air spring with bellows has due to the stiffness change a virtually constant body natural frequency in the entire load range.

To conclude, Figure 6 shows the influence of different suspension systems on the natural frequency, and thus also indirectly on the comfort with rising loads. The reason for the influence of the body natural frequency on comfort lies in the different resonance ranges of different organs in the human body and the consequence that an excitation of human body parts with their natural frequency impairs well-being. This is why a body natural frequency, below the resonant frequencies of the human body, that is as independent as possible of the load is to be ensured.

However, Figure 6 also clearly indicates that an approximately constant natural frequency with rising load is only present in the case of air springs. With steel springs, the natural frequency drops due to the constant spring stiffness.

Figure 6: Comparison of different suspension systems depending on payload
1 Steel spring,
2 Air spring,
3 Hydropneumatic spring.

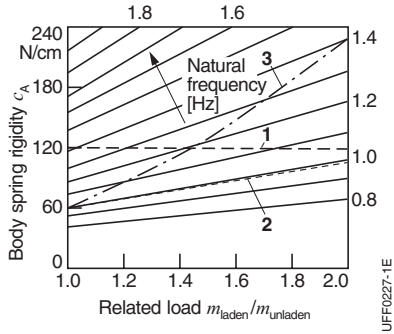
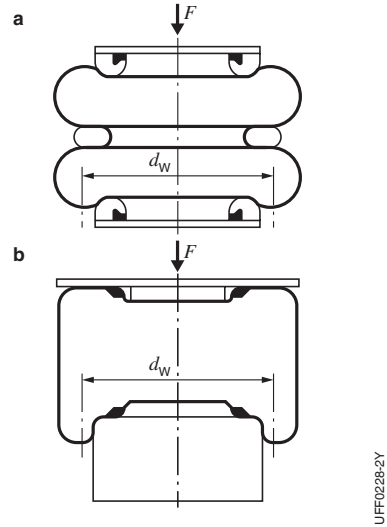


Figure 7: Gas springs with bellows
a) Bellows,
b) U-type bellows.
F Force,
d_w Effective diameter of air spring.



Springs with bellows
Gas springs with bellows with pneumatic ride-height control are suspension systems with constant gas volume (see above), whereby these are in turn divided into two categories. These are firstly gas springs with bellows and, secondly, U-type bellows gas springs (Figure 7) which, in a similar way to pneumatic tires, consist of rubber material reinforced by woven textiles. On these systems, the ride-height control is implemented by pumping gas into or releasing gas from the spring with the suspension volume remaining constant as a general principle. The effective surface area of the air springs (and thus the gradient of the return force) that is affected by the overpressure is not usually constant, rather it changes throughout the stroke. This enables specific influence on the load-bearing capacity by designing the contours of the rolling piston of the U-type bellows gas springs (and thus a change in the effective surface area A in Equation 3 via the range of the spring) accordingly. The effective surface area A of the air spring can be determined via the effective diameter. With integration of an additional volume (increase of h_{th}, see Figure 5), a less progressive and flatter-tending characteristic curve can be achieved.

Hydropneumatic springs
In accordance with the considerations above, hydropneumatic springs (Figure 8) with integrated ride-height control are gas springs with constant gas weight (see above), whereby the power flux is routed not only through a gas but additionally through a fluid or oil. Here, the fluid and gas are separated by an impermeable rubber membrane. It is only when the fluid or oil is placed in between that a wear-resistant and low-friction seal is achieved between the piston and cylinder.

Another advantage of this system lies in the possibility to integrate hydraulic damping in the suspension design element. A disadvantage, on the other hand, is the dependency of the natural frequency on the load (influence on comfort). The reason for this is the pumping in and out of the fluid or oil at a constant gas mass that is required for ride-height control. The load-sensitive volume change of the gas leads to a shift in the spring stiffness in such a way that with rising load there is a fundamental increase in natural frequency (see Figure 6).

Figure 8: Hydropneumatic spring
1 Suspension ball, 2 Membrane,
3 Connection to ride-height controller,
4 Piston, 5 Screw cap,
6 Suspension cylinder, 7 Cup seal,
8 Shock-absorber valve, 9 Rebound stage,
10 Bypass, 11 Compression stage.

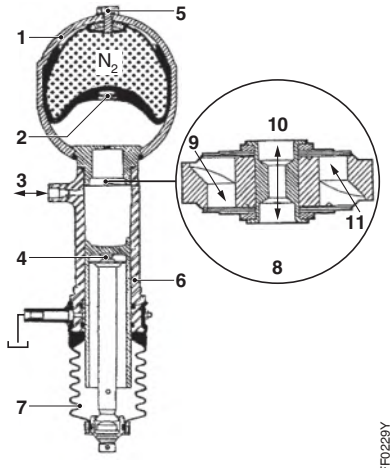
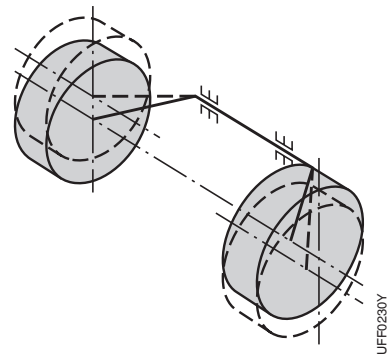


Figure 9: Functioning principle of stabilizer spring



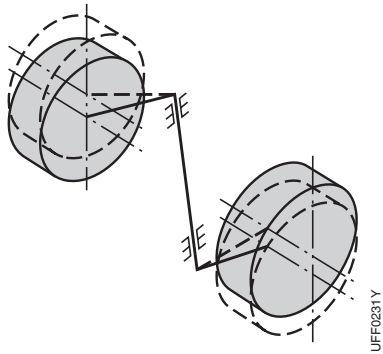
Stabilizers
The suspension systems described above are deployed primarily for the vertical suspension of a vehicle. For the roll suspension, on the other hand, additional passive or active stabilizer springs (under certain circumstances with additional roll damping) are used alongside classical body springs. A diagram of the principle is provided in Figure 9. In the event of a rolling movement of the body, i.e., spring compression of the wheels of one axle in opposite directions, the stabilizer is torsioned and delivers an aligning torque around the roll axis. In the case of in-phase vertical motions of one axle, on the other hand, this has no effect. If the proportions of the rolling moment braced by the stabilizers on the front and rear axles have a different ratio to the proportions braced by the body springs, not only is the roll angle reduced but the breakdown of the differences in wheel load of an axle on cornering and thus the self-steering effect are also influenced.

On a vehicle with a corresponding stabilizer configuration, this can shift the driving characteristics towards understeer (increase in the roll rigidity on the front axle or reduction in the roll rigidity on the rear axle) or oversteer with all other parameters unchanged. In the case of active stabilizers, the stabilizer force can also be actively influenced and adapted to suit the driving status. This enables a reduction in for example copy effects on one axle (due to decoupling between the right-hand and left-hand sides) when the vehicle is driven straight ahead but it also increases the driving dynamics on cornering by minimizing the inclination of the body. Here, the stabilizers do not influence the vertical vibration characteristics of the vehicle.

Compensating springs
Compensating springs have the opposite effect to that of the stabilizer (Figure 10). As a pure stroke element, they have no effect during rolling motions of the body.

The compensating spring was used in the past on axle designs where the wheel suspension kinematics required wheel load differences that were as low as possible to suppress the “resting effect” (stronger rebound of the wheel on the inside of the curve compared to the spring compression of the wheel on the outside of the curve when cornering). The stiffness of the body springs and this braced proportion of the rolling moment on the axle under examination could then be reduced accordingly. Compensating springs are no longer in modern passenger car wheel suspensions.

Figure 10: Functioning principle of compensating spring



Suspension systems

Increasing customer requirements (comfort and driving dynamics) with regard to passenger cars and the strongly fluctuating load states of commercial vehicles mean that exclusive deployment of conventional steel springs is often insufficient. In such cases, either partially loaded or fully loaded suspension systems are used.

The integration of the additional functions of partially loaded or fully loaded systems enables increases in both comfort and driving dynamics (for example transverse locking of the springs of one axle to enhance stability on cornering).

Partially loaded systems

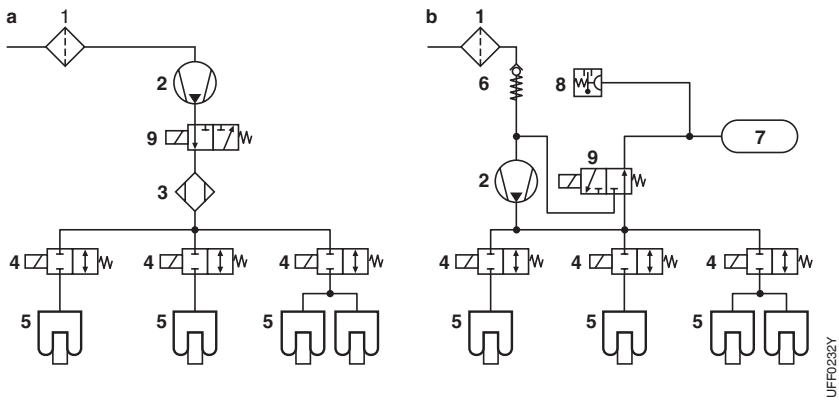
These systems are characterized by the fact that the forces to be braced by the suspension system are divided between steel and air springs according to a specified ratio.

In the case of soft body springs (to enhance ride comfort), wide spring ranges occur for example when a vehicle is loaded. In order to prevent the vehicle body from being lowered too much, ride-height control systems with air springs or hydropneumatic springs are used. Here, sensors determine the ride height and provide this information to the control system. By pumping in or releasing air (air springs with bellows) or oil (hydropneumatic springs), the ride height can be then adapted in line with requirements. Depending on the vehicle segment (passenger cars or commercial vehicles), ride-height control systems offer different additional functions.

In the passenger car, for example, a speed-dependent ride-height control of the body in order to save fuel is possible. The adjustable ride height can also be used on poor-quality road surfaces to enhance the vehicle's capability to handle rough terrain.

In the case of commercial vehicles, on the other hand, ride-height control enables variable adaptation of the loading area to different loading ramps. Other functions can also be implemented by networking with other systems. These include, for example, an automatic ride-height increase if the lifting axle is raised, lowering on exceeding the maximum axle load, or brief raising of the lifting axle to increase the wheel load on the driven axle.

Figure 11: System architectures of ride-height control, fully loaded systems
a) Open system, b) Closed system.
1 Filter, 2 Compressor, 3 Drier, 4 2/2-directional-control valve, 5 Bellows, 6 Non-return valve, 7 Pressure container, 8 Pressure switch, 9 3/2-directional-control valve.



Fully loaded suspension systems

In the case of fully loaded systems, as opposed to partially loaded solutions, the task of suspension is assumed solely by the gas springs and the helical springs are completely eliminated. Depending on the available hardware and control strategy, the ride-height control can be on selected axles or on all axles. Embedding the control system architecture in the global vehicle control architecture ensures that a negative mutual influence of the axles is excluded, thus preventing, for example, the vehicle body from slanting.

As a general principle, the ride-height control for fully loaded systems can be designed in the form of an open or closed system. In the case of the open system, a compressor draws air from the atmosphere and provides it in compressed form to the air springs if required. The increase in pressure raises the vehicle body. To lower the ride height, air is expelled into the environment, thus lowering the pressure in the springs again. Although this system involves relatively low overhead in construction and has a simple control system, it has the decisive disadvantage that high compressor output has to be provided for the short periods of control operation, which is accompanied by corresponding increased consumption. Moreover, an air drier is required and an acoustic load can be expected when air is drawn in and expelled.

The closed system draws air from a pressure accumulator of the suspension system and feeds this directly to the air springs. When the ride height is lowered, the compressed air is returned to the pressure tank. Although the compressor output requirements for this system are lower and the air drier can be eliminated because the working medium is already dry, other components (accumulator, pressure switch, non-return valve, return line etc.) are required, fundamentally increasing the overhead in construction compared to an open system.

The system architectures of the open and closed ride-height control systems are compared in Figure 11. This clearly illustrates the greater complexity of the closed system.

Active spring systems

Compared to passive systems, active vehicle chassis systems enable optimal adaptation of spring and damper forces to all driving states and road irregularities. With the aid of external sources of energy, forces are generated that stabilize both the axles and the vehicle body. In the meantime, a number of different system architectures have been developed. These differ above all with regard to the overhead involved (design envelope and costs), the energy requirement, and the quality of control operations. A few of these systems are presented briefly below.

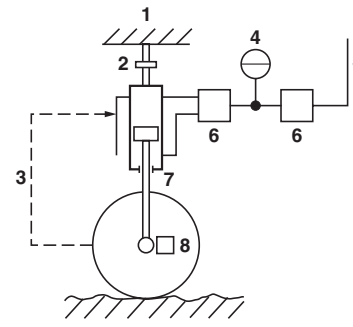
Systems with hydraulic cylinders

With this version, the body motion and body position are regulated with quickly adjustable hydraulic cylinders (Figure 12), whereby different items of sensor information (wheel load, range of spring, acceleration etc.) are used as input variables for the control operation.

The control system achieves virtually constant wheel loads while maintaining a constant mean ride height. The static wheel load in this case is carried by steel springs or hydropneumatic springs.

Figure 12: Active chassis system with hydraulic cylinder

1 Vehicle body, 2 Wheel-load sensor,
3 Travel sensor, 4 Accumulator,
5 Pump circuit, 6 Servo valve,
7 Positioning cylinder, 8 Acceleration sensor.



Systems with hydropneumatic suspension system
To stabilize the vehicle, a hydropneumatic suspension system uses specific oil regulation. This is done by pumping hydraulic oil into the suspension struts or draining it off the suspension struts (Figure 13). To limit the energy absorption, the control strategy of this system consists of control operations to eliminate longwave irregularities (low-frequency excitations). In the case of higher-frequency proportions, a gas volume takes effect in the proximity of the suspension strut. Here, the shock absorber is essentially geared to the wheel movements.

Version with spring-mounted point adjustment
In this system, the vehicle body is held horizontally in the low-frequency range in that a conventional helical spring is designed as variable at its mounting point (either against the vehicle body or in relation to the axle, Figure 14). The base is raised for spring compression and lowered for spring rebound. Control operations are continuous by means of a fluid pump and proportioning valves. The helical spring, however, must be longer compared to its original design. In this type of system, the shock absorbers can be fitted with constant adjustment parameters and, above all, they can be matched to the wheel dampers.

Electromagnetic system
In the case of an electromagnetic active chassis system, linear or rotary electromagnetic motors are fitted on each wheel. These are able to actively compensate for the road irregularities. The linear motors are supplied with electrical energy by power amplifiers, whereby as a general principle force control or path control can be implemented. Cross-axle system networking also enables compensation

Figure 13: Hydropneumatic system
1 Vehicle body, 2 Travel sensor,
3 Accumulator, 4 Pump circuit,
5 Acceleration sensor, 6 Throttle,
7 Proportioning valve,
8 Shock-absorber piston fitted with valves.

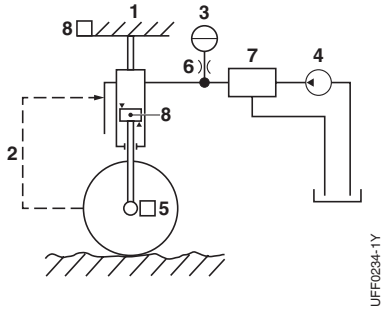
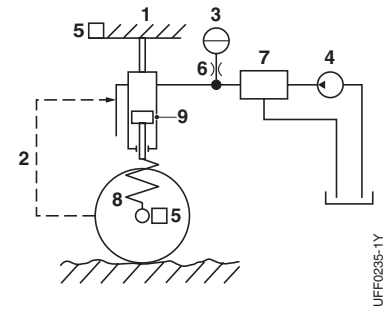


Figure 14: Spring-mounted point adjustment
1 Vehicle body, 2 Travel sensor,
3 Accumulator, 4 Pump circuit (oil),
5 Acceleration sensor, 6 Throttle,
7 Proportioning valve,
8 Body spring (coil spring),
9 Spring-mounted point adjuster.



of rolling and pitching vibrations. In the solution that is currently implemented, the static wheel loads are absorbed via torsion springs at the wheels to limit the electrical energy requirement. A passive shock absorber is also used.

The advantages of the electromagnetic solution lie above all in the high adjustment speed. Despite the low outer dimensions, the electric motors have sufficient power output to ensure driving safety in all driving situations. In comparison with conventional systems, the disadvantages lie in the greater weight and in the increased costs, as an additional, non-electronic damping system is required for safety reasons. As a general principle, however, the electromagnetic motor can also be operated as a generator, which means that damping can also be performed electrically and thereby energy can be recuperated. This enables a lowering of the power requirement of the overall system, which is specified for standard road surfaces as less than 1 kW.

References

[1] B. Heissing, M. Ersoy (Editors): Fahrwerkhandbuch. 1st Edition, Vieweg Verlag, 2007.

Shock absorbers and vibration absorbers

Shock absorbers

The masses of the vehicle body and wheels connected by the body springs form a system that can vibrate and is excited by irregularities on the roadway and the dynamic movements of the vehicle. Shock absorbers are required to dampen the vibrating system. Nowadays, it is almost exclusively hydraulic telescopic shock absorbers that are used in motor vehicles as body shock absorbers; these convert the kinetic energy of the body and wheel vibrations into heat. The shock absorbers are configured taking account of the partially conflicting requirements for comfort (minimization of body accelerations) and driving safety (minimization of wheel load fluctuations).

Basic principles of hydraulic telescopic shock absorbers

The damper effect of hydraulic telescopic shock absorbers (Figure 1) is based on the movement subject to flow resistance of a shock-absorber piston equipped with throttle elements (damper valves) within an oil-filled working cylinder. In the process, mechanical work is converted into heat that is released into the environment via the shock-absorber surface. The pressure difference Δp between the two working chambers and the effective surfaces caused by the throttle elements on both sides of the shock-absorber piston generates the resulting damping force F_D when the shock absorber moves in or out. The area on which the pressure prevailing in each working chamber is applied corresponds to an annular surface A_{KR} for the working chamber through which the piston rod of the shock absorber runs (see working chamber 1 in Figure 1). The outer diameter of the annular surface corresponds to the diameter D of the shock-absorber piston; the inner diameter corresponds to the diameter d of the piston rod. The following applies:

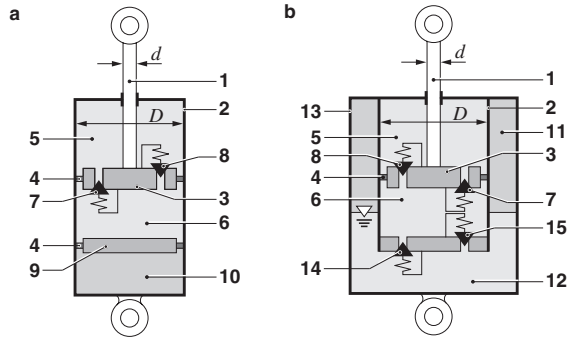
$$A_{KR} = \frac{\pi}{4}(D^2 - d^2).$$

In the other working chamber (see working chamber 2 in Figure 1), the effective area corresponds to the piston surface A_K which results from the diameter D of the shock-absorber piston.

$$A_K = \frac{\pi}{4}D^2.$$

Figure 1: Structure of single-tube and twin-tube shock absorbers

a) Single-tube shock absorber, b) Twin-tube shock absorber.
1 Piston rod, 2 Working cylinder, 3 Shock-absorber piston, 4 Piston seal, 5 Working chamber 1, 6 Working chamber 2, 7 Piston valve (rebound stage valve), 8 Piston valve (compression stage valve), 9 Dividing piston, 10 Gas volume, 11 Part of compensating chamber, filled with gas, 12 Part of compensating chamber, filled with shock-absorber oil, 13 Outer tube, 14 Floor valve (compression-stage valve), 15 Floor valve (rebound-stage valve).
D Inner diameter of working cylinder and diameter of shock-absorber piston,
d Diameter of piston rod.



When there is a movement of the shock-absorber piston (i.e, an inward or outward movement of the shock absorber), the changes to the volumes in the two working chambers lead to a flowing movement of the incompressible damping fluid between the working chambers of the shock absorber (in the case of twin-tube shock absorbers, additionally between a working chamber and the compensating chamber). The oil volume flows in each case must pass the corresponding valves.

The oil-volume flows through the relevant valves result from the geometry of the shock absorber and the speed of inward or outward movement $\dot{z}$. The following applies to the volume flow $\dot{Q}_1$ between the two working chambers:

$$\dot{Q}_1 = \frac{\pi}{4}(D^2 - d^2)\,\dot{z}.$$

The inward or outward movement of the piston rod on hydraulic telescopic shock absorbers leads to a variable total volume of the working chambers that depends on the retracting or extending state. The incompressibility of the shock-absorber oil means that a possibility to compensate for the oil volume displaced or released by the piston rod is required. The following applies to the volume flow $\dot{Q}_2$ of this compensating volume:

$$\dot{Q}_2 = \frac{\pi}{4}\,d^2\,\dot{z}.$$

The volume flow $\dot{Q}$ through a valve is linked via the throughflow characteristics of each valve to the prevailing pressure difference Δp. The throughflow characteristics of a valve result from the joint effect of the throttle geometry (for example bore diameter of the flow-through channel) and any spring load (i.e., pressure-dependent variation of the discharge opening, Figure 1). The throughflow characteristics can be adapted to the needs of each situation by configuring and coordinating these parameters. The characteristics of the valves are to be designed in such a way that no cavitation whatsoever occurs (formation and implosion of gas bubbles in the working medium due to static pressure fluctuations in the range of the vapor pressure of the working medium) inside the shock absorber. Cavitation leads to acoustic problems and also to damage – ultimately to failure of a shock absorber. The cavitation tendency can also be effectively counteracted by pressurizing the shock-absorber oil.

Types of hydraulic telescopic shock absorbers

Single-tube shock absorbers

To balance out the retracting or extending piston-rod volume, single-tube shock absorbers have an enclosed gas volume which is separated from the working chambers filled with shock-absorber oil with the help of a moving dividing piston (Figure 1a). In the compression phase (moving in) of the shock absorber, the gas volume is compressed according to the volume flow $\dot{Q}_2$; in the rebound phase (moving out), it is relaxed according to the volume flow $\dot{Q}_2$. As a rule, the pressure of the gas volume is between 25 and 35 bar, which means that the maximum occurring retraction forces (compression-stage force) can be absorbed without creating a vacuum at the piston-road seal. The working volume flow $\dot{Q}_1$ flows through each of the piston valves. On moving in, it flows though the compression-stage valve and on moving out through the rebound-stage valve.

The high gas pressure means that the tendency for cavitation to occur is low in the case of the single-tube shock absorber. The heat that is generated can be released into the environment directly via the outer surface of the working cylinder. The advantages of lean design, the low weight and discretionary installation position of the single-tube shock absorber are offset by the great length, increased friction due to the high internal pressure , and high requirements with regard to sealing the piston rod and gas volume. Furthermore, the single-tube shock absorber is, on account of its more filigree design, not suitable for absorbing considerable lateral forces and bending moments.

Twin-tube shock absorbers

Twin-tube shock absorbers have a compensating chamber resulting from the arrangement of an outer tube around the working cylinder (Figure 1b). The com-

pensating chamber balances out the retracting or extending piston-rod volume. To achieve this, it is connected via bottom valves to the lower working chamber of the shock absorber. The compensating chamber is partly filled with shock-absorber oil and partly with a gas (as a rule, dried air or even nitrogen). The gas volume is at atmospheric pressure or slight overpressure (6 to 8 bar). An overpressure here only serves the purpose of reducing the cavitation tendency; compression-stage-force absorption is not necessary. The piston and bottom valves must be coordinated is such a way that no cavitation occurs. In the compression stage, i.e., when the shock absorber is moving in, the damping work is therefore done at the corresponding bottom valve (compression-stage valve) through which the volume flow $\dot{Q}_2$ flows. On the other hand, the oil volume flow $\dot{Q}_1$ from the lower working chamber into the upper working chamber can only flow through the compression-stage valve in the shock-absorber piston with low flow resistance. This prevents a radical drop in pressure in the upper working chamber. In contrast, in the rebound stage, i.e., when the shock absorber is moving out, the damping work is essentially done by the volume flow $\dot{Q}_1$ from the upper working chamber into the lower working chamber at the corresponding piston valve (rebound-stage valve). The bottom valve only balances out the extending piston-rod volume in that shock-absorber oil flows virtually without resistance from the compensating chamber into the lower working chamber (volume flow $\dot{Q}_2$).

The compensating chamber means that twin-tube shock absorbers have poorer heat dissipation in comparison with single-tube shock absorbers. Furthermore, the installation position of twin-tube shock absorbers is restricted, as it must be ensured at all times that there is compensating fluid at the bottom valves. Advantages compared to the single-tube shock absorber are the lower shock-absorber length and the softer responsiveness due to the seals being subject to less load with correspondingly lower requirements. Furthermore, thanks to its twin-walled structure, the twin-tube shock absorber, with a suitably more solid piston-rod and outer-tube design, can absorb lateral forces and bending moments, meaning that it can be used as a wheel-control element (spring-strut or shock-absorber-strut independent wheel suspension). In the area of passenger cars, the twin-tube shock absorber has prevailed as the standard shock absorber – also due to its lower costs.

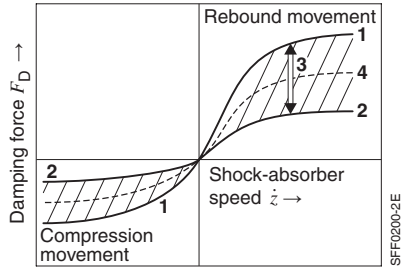
Adjustable shock absorbers

The conflict of objectives with regard to the coordination of body shock absorbers between ride comfort and driving safety can be mitigated using adaptive or semi-active shock absorbers. In comparison with passive shock absorbers with fixed shock-absorber characteristics (i.e., defined force-speed characteristics, cf. section “Damping characteristics”), adaptive shock absorbers provide the possibility of discrete to infinitely variable adjustment of the damping characteristics (Figure 2).

Alongside manual adjustment of the shock absorbers (for example soft damping in the comfort mode or hard damping in the sport mode), adjustable shock absorbers can also be activated automatically depending on the driving state in each case (see section “Damping control”). The shock absorbers are as a rule adjusted by the driver by way of controls

Figure 2: Adjustable shock-absorber characteristic curves

1 Upper limit of adjustment range (maximum damping hardness, e.g., sport mode),
2 Lower limit of adjustment range (minimum damping hardness, e.g., comfort mode),
3 Adjustment range of shock absorber,
4 Characteristic curve of a passive shock absorber.



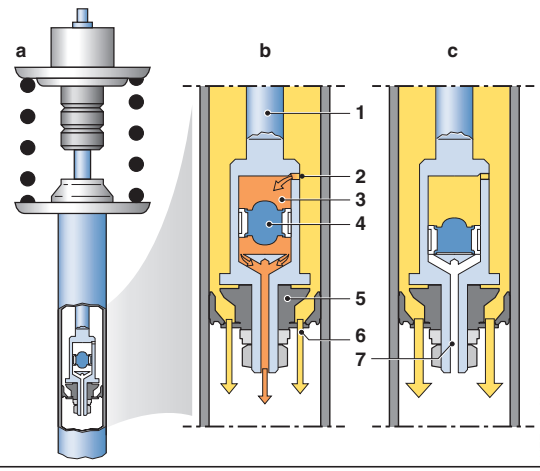
in the instrument panel. Adjustability directly at the shock absorber itself is encountered primarily in motor sport and in sporty retrofit shock absorbers.

Adaptive hydraulic absorbers
In the case of adaptive or semi-active shock absorbers of conventional design, the adjustability of the damping characteristics is implemented by means of adjustable valves, bypass boreholes that can be activated (located on the outside or inside), or with the help of double pistons [1]. As a rule, activation is electronic. Major system features are the adjustment times that can be achieved, the spread of the adjustment range, as well as the number of definable shock-absorber characteristics. It must be mentioned here that the adjustment times should not be shortened arbitrarily in the interests of functionality because changing the flow conditions in the shock absorber too quickly can give rise to noise emissions and possibly wear. Whereas systems of the first generation only permitted adjustment between a few characteristic curves, today's adaptive shock absorbers can usually be set to a large number of characteristics [2]. Newer systems usually even have infinitely variable adjustment between a minimum and a maximum damping-force characteristic (Figure 2).

Rheological shock-absorber systems
The adjustability of the damping characteristics in the case of rheological shock-absorber systems is based on the change in the flow properties of the working medium that is used. Magneto-rheological fluids that change their viscosity under the influence of a magnetic field are used here instead of the usual mineral oils. Electro-rheological fluids have similar properties under the effect of an electric field, but the effect is utilizable to a much lesser extent, which is why primarily magneto-rheological methods have prevailed. The viscosity of the working medium has a direct influence on the flow resistance through the valves. If, for example, creating a magnetic field increases the viscosity of a magneto-rheological working medium, the flow resistance through the valves increases.

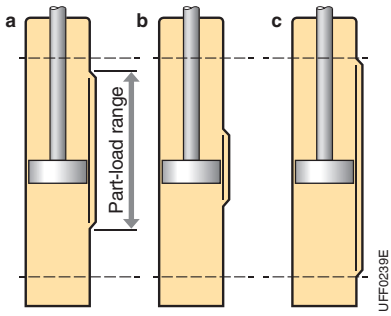
Rheological shock-absorber systems not only provide the possibility for infinitely variable adjustment of the damping characteristics but also the implementation of very short adjustment times [2].

Figure 3: Amplitude-selective shock absorber
a) Suspension/shock-absorber system,
b) Damping, normal driving style,
c) Control piston at end stop blocks the bypass channel, resulting in stronger damping effect with a dynamic driving style and in curves (bends).
1 Piston rod,
2 Restriction bore,
3 Working chamber,
4 Control piston,
5 Piston,
6 (Main) oil volume flow through piston valve,
7 Bypass channel.



Passively adaptive systems
Aside from the mentioned shock-absorber designs with different options for characteristic-curve adaptation by the driver (electronically at the touch of a button or mechanically by hand) or by control electronics, there are also types which are fashioned in such a way that the shock-absorber characteristic is adapted within limits to different driving-dynamic influences. Amplitude-selective, stroke-dependent, and frequency-selective damping are mentioned here by way of example. In the case of amplitude-selective damping a bypass channel with an implemented control piston is usually provided in the shock-absorber piston (Figure 3). In a limited amplitude range the oil flows in part through this bypass and displaces the control piston, increasing the flow cross-section and thus reducing the damping forces. When the control piston reaches its structurally predetermined end stop, the oil now flows only through the piston valve and the damping forces increase. The freely moving section of the control piston determines the amplitude up to which the damping forces are reduced, usually a few millimeters. The result of this is that (depending on the load) smaller road irregularities are damped much less than for example than pitching or rolling of the vehicle, which provides for good comfort and at the same time good driving dynamics.

Figure 4: Stroke-dependent shock absorber
a) Two-stage damping: damping dependent on load/stroke,
b) Two-stage damping: damping dependent on dynamic load (with ride-height control),
c) Hydraulic end-position damping for rebound/compression stage.



Stroke-dependent shock absorbers
In stroke-dependent shock absorbers the operating principle of the partially enlarged flow cross-section is effected by bypass grooves over a predetermined length in the shock-absorber tube (Figure 4). The total flow cross-section is enlarged in the stroke ranges in which the piston passes over these bypass grooves, with the result that the damping forces are lower in this range. The damping force increases in the case of larger strokes outside the bypass-groove range (e.g., towards the end stop). Because of the fixed position of the grooves this system is dependent on the stroke range and the piston position in the shock absorber and less so on the oscillation amplitude.

Frequency-selective shock absorbers
In the case of frequency-selective shock absorbers the cross-section is altered by spring-loaded, oscillatory valves which according to the tunning at predetermined frequencies (e.g. body natural frequency) reduce the flow cross-section and consequently increase the damping forces. Less damping occurs outside the defined frequency range.

Damping characteristics
Damping force is a function of the speed of the inward or outward movement of the shock absorber, whereby the direction of force is opposed to the direction of speed at all times. It is generally applicable that the damping force F_D and the speed $\dot{z}$ are linked via the damping constant k_D and the damping exponent n. The following applies:

$$F_D = -\text{sign}(\dot{z}) \cdot k_D \cdot |\dot{z}|^n .$$

The damping constant and damping exponent are essentially dependent on the design of the shock absorber (valve characteristics, geometry). With the corresponding configuration of the individual parameters, progressive to degressive damping characteristic curves can be

created. Varying the bore diameter in the shock-absorber piston influences primarily the progression (tube flow), changing the spring load influences first and foremost the degressivity of the characteristic curve (increase in cross-section with increasing pressure difference). Modern body shock absorbers have mainly degressive characteristics. This achieves a high damping effect at low excitation speeds as well as a limitation of the maximum damping forces.

Damping characteristic curves are usually determined with the help of mechanical or servo-hydraulic testing units. A sinusoidal path excitation of constant amplitude and variable frequency or constant frequency and variable amplitude results in various maximum speeds of the inward or outward movement. The recorded path and force signals can be applied in a force-path diagram (work diagram) (Figure 5a). The force-speed characteristic curve of the shock absorber (damping characteristic curve) can be derived from the work diagram by transferring the maximum force and speed values (Figure 5b).

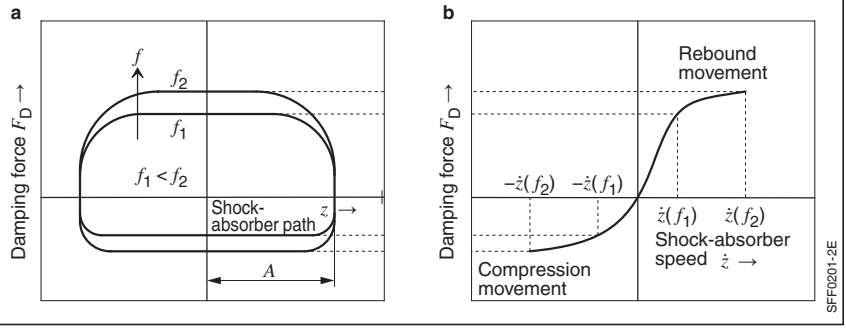
It is mainly for reasons related to comfort that the configurations of the rebound and compression stages differ. The damping forces generated in the rebound stages are for the most part more than twice as high as the correspondingly generated forces during compression (i.e., in the compression stage; ratio for passenger cars is 1:2 to approximately 1:3, for trucks up to 1:9) of the shock absorber (Figure 5). This limits the impact forces on the vehicle body during the compression phase (comfort) and simultaneously ensures the system is strongly dampened (system relaxation) in the rebound phase. In the event of an asymmetrical distribution of the required total shock-absorber work in favor of a high rebound stage there is the danger of the shock absorber contracting with soft springs.

Damping control

In conjunction with electronically adjustable shock absorbers, damping control systems are being used to an increasing degree nowadays. The major component parts of such damping-control systems are the adaptive shock absorbers and sensors (e.g., acceleration sensors on the wheel and body mass) and intelligent algorithms and control strategies. With the help of the sensors and algorithms, the current driving state is continuously determined and evaluated. In accordance with the stored control strategies, this enables the control system to adapt the shock-absorber characteristics to each driving state by activating the shock ab-

Figure 5: Damping characteristics

a) Work diagram (force-path diagram), b) Damping characteristic curve (force/speed diagram).
f Variable excitation frequency, f_1 Excitation frequency 1, f_2 Excitation frequency 2,
$\dot{z}(f_1)$ Maximum shock-absorber speed at f_1, $\dot{z}(f_2)$ Maximum shock-absorber speed at f_2,
A Constant excitation amplitude.



sorbers, thus, for example, influencing and optimizing ride comfort or driving safety. Such systems are often also called semi-active because no active spring adjustment actually occurs, the wheel movement however can be well defined via the adjustable shock absorber within limits.

Control strategies

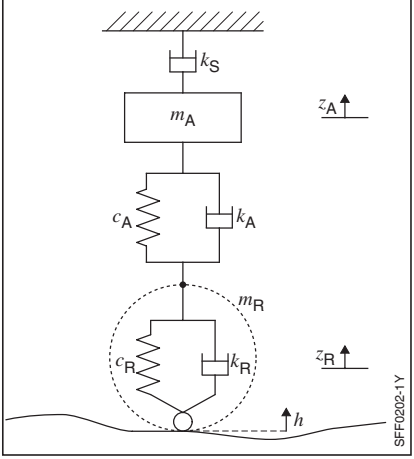
Threshold-value strategy

Threshold-value controllers compare one or a number of relevant driving-state variables (e.g., body acceleration, steering angle) with the corresponding limit values and initiate defined measures if these values are exceeded or not reached. The damping forces are usually influenced axle by axle simultaneously in the rebound and compression directions. The main focus of threshold-value controllers is on increasing comfort while simultaneously maintaining driving safety.

Alongside purely influencing the vertical vibration characteristics, induced vehicle-body movements can also be optimized. For example, steering angle monitoring can lead to a reduction in the dynamic rolling or hardening the damping depending on the brake pressure can reduce pitching motion caused by braking.

Figure 6: Theoretical principles of the skyhook approach

k_S Damping constant of skyhook shock absorber,
m_A Body mass,
z_A Vertical body motion,
c_A Spring stiffness of body spring,
k_A Damping constant of body shock absorber,
m_R Wheel mass (unsprung mass),
z_R Vertical wheel movement,
c_R Vertical tire spring stiffness,
k_R Damping constant of tire,
h Vertical road excitation.



Skyhook

The skyhook control strategy aims to keep the vehicle body calm independently of current driving state and road conditions. This is intended above all to increase ride comfort. In contrast to the threshold-value strategy, the skyhook control strategy regulates the damping characteristics at each individual wheel. The basic principle is to decouple the movement of the vehicle body from the road excitation. To achieve this, it is imagined that the vehicle body is connected by means of a shock absorber to the sky (Figure 6). The damping force F_{DS} of the skyhook shock absorber results from the link of the body speed $\dot{z}_A$ and the damping constant k_S of the imaginary sky shock absorber:

$$F_{DS} = k_S \dot{z}_A .$$

For the conventional vibration system, on the other hand, the damping force F_D would result from the link of the damping constant k_A of the body shock absorber and the difference between the vertical body speed $\dot{z}_A$ and the vertical wheel speed $\dot{z}_R$:

$$F_D = k_A(\dot{z}_A - \dot{z}_R) ,$$

In order to brace the vehicle body against the sky, in the real implementation the additional portion of force F_{DS} of the sky shock absorber must be applied by the body shock absorber. The proportional damping factor k_{AS} this requires is calculated to:

$$k_{AS} = \frac{k_S \dot{z}_A}{\dot{z}_A - \dot{z}_R} .$$

As an adaptive (semi-active) shock absorber is only able to extract energy from the system in the form of heat, but is un-

able to supply heat to the system, a case distinction is required [1], [2]. The following applies:

$$F_{Dtot} = \left(k_S \frac{\dot{z}_A}{\dot{z}_A - \dot{z}_R} + k_A \right) \cdot (\dot{z}_A - \dot{z}_R)$$

where $\dot{z}_A (\dot{z}_A - \dot{z}_R) \geq 0$, and

$$F_{Dtot} = k_A (\dot{z}_A - \dot{z}_R)$$

where $\dot{z}_A (\dot{z}_A - \dot{z}_R) < 0$.

Depending on the amount and direction of the vehicle-body speed and the shock-absorber movement (rebound or compression movement), the control strategy for the skyhook approach shown in Figure 7 is pursued for comfortable damping of the vehicle body. However, specific damping of rolling, pitching, and wheel vibrations is not taken into account by this approach. These movements are also highly relevant with regard to ride comfort and driving safety, which means that the skyhook controller is generally overridden by other controllers.

Groundhook

A groundhook controller aims to improve driving safety by reducing the wheel load fluctuations. In much the same way as the considerations for the skyhook strategy, it is imagined that the wheel is connected

Figure 8: Theoretical principles of the groundhook approach

m_A Body mass,
z_A Vertical body motion,
c_A Spring stiffness of body spring,
k_A Damping constant of body shock absorber,
m_R Wheel mass (unsprung mass),
z_R Vertical wheel movement,
c_R Vertical tire spring stiffness,
k_R Damping constant of tire,
h Vertical road excitation,
k_G Damping constant of groundhook shock absorber.

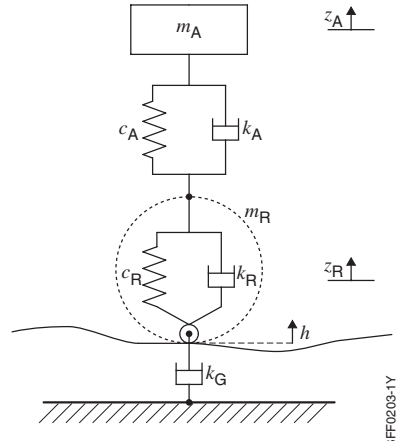
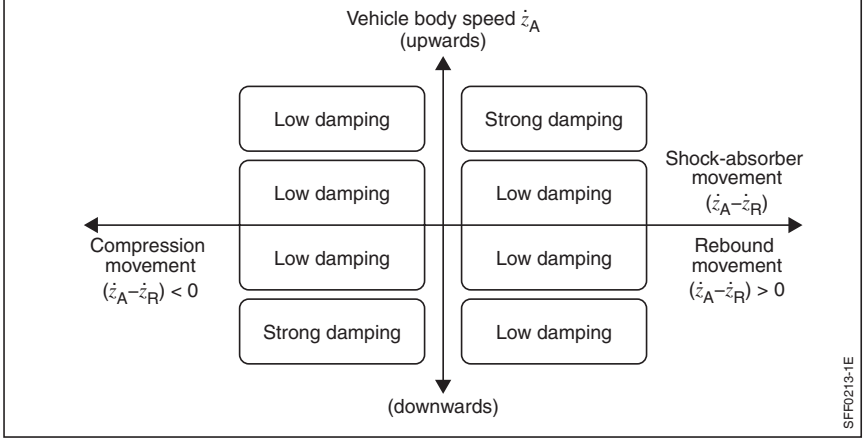


Figure 7: Control strategy of skyhook control (case distinction)



by a shock absorber to the roadway (Figure 8) and a proportional damping factor k_{AG} is derived. The following applies in the same way as the derivations for the skyhook controller:

$$k_{AG} = k_G \frac{\dot{z}_R - \dot{h}}{\dot{z}_R - \dot{z}_A}$$

where:
$\dot{z}_A$ Vertical body speed,
$\dot{z}_R$ Vertical wheel speed,
$\dot{h}$ Vertical excitation speed,
k_G Damping constant of the groundhook shock absorber.

Also for the groundhook controller, a case distinction is made depending on the directions of movement of the wheel and vehicle body. This case distinction takes place on the basis of the term:

$$(\dot{z}_R - \dot{h})\ (\dot{z}_R - \dot{z}_A)\ .$$

Huang's algorithm
The aim of Huang's algorithm is shock-absorber control in favor of as low an effective body acceleration as possible as a comfort criterion [3]. With high body acceleration and rectified damping force, damping is set as high as possible. If damping force and body acceleration are opposed, damping is set as low as possible.

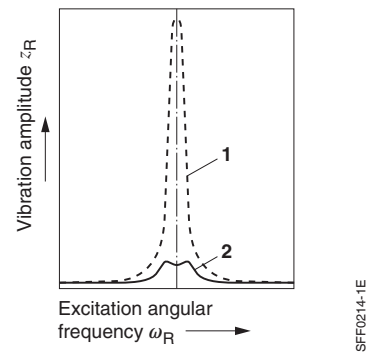
$\ddot{z}_A(\ddot{z}_A - \ddot{z}_R) > 0 \rightarrow$ hard characteristic,
$\ddot{z}_A(\ddot{z}_A - \ddot{z}_R) \leq 0 \rightarrow$ soft characteristic.

Vibration absorbers

To specifically influence the vibration properties of the vibration system consisting of the wheel and vehicle-body mass, vibration absorbers (see Vibrations absorption) are deployed in some cases in the area of the chassis.

Depending on the configuration and arrangement of the vibration absorber, the comfort, acoustics, or driving safety can be influenced. A distinction is made between passive and active vibration absorbers. A passive vibration absorber is a mass attached to the chassis by sprung and damped mountings (Figure 10). The absorbing effect is created by the corresponding mass forces and in the case of passive absorbers is restricted to a certain frequency range. The effective range can be enlarged using an active absorber with an actuator that can be activated.

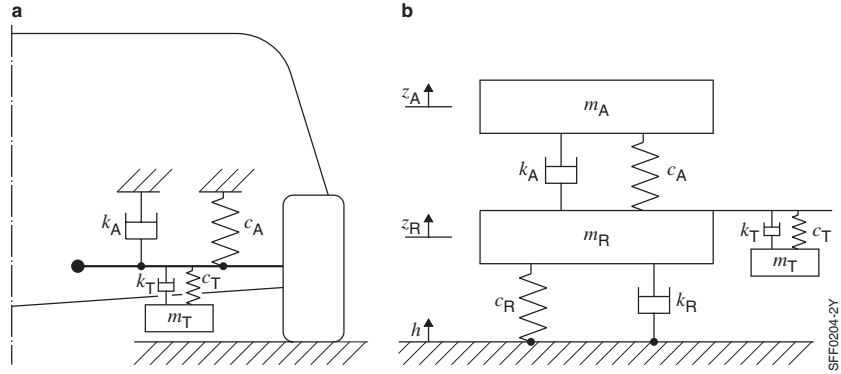
Figure 9: Vibration amplitude of a wheel movement as a function of the excitation angular frequency with and without vibration absorbers
1 Progression without vibration absorber,
2 Progression with vibration absorber.



On excitation of the vibration system, the vibrations and oscillations of the main system are taken over by the appropriately coordinated vibration absorber, i.e., the main system only vibrates very slightly whereas the absorber absorbs a large portion of the energy. Figure 9 shows an example of the progression of the vibration amplitude of a wheel movement with and without vibration absorbers. With the deployment of a vibration absorber geared to the frequency range of the natural wheel frequency, a significant drop in the vibration amplitude can be observed in the corresponding frequency range. However, vibration absorbers have the disadvantage of a (depending on the frequency range to be extinguished) high additional mass.

References
[1] B. Heissing, M. Ersoy (Editors): Fahrwerkhandbuch. 4th Ed., Verlag Springer Vieweg, 2013.
[2] L. Eckstein: Aktive Fahrzeugsicherheit. ika/fka 2010.
[3] P. Zeller (Editor): Handbuch Fahrzeugakustik – Grundlagen, Auslegung, Berechnung, Versuch. 3rd Edition, Verlag Springer Vieweg, 2018.

Figure 10: Vibration absorber
a) Vibration absorber in chassis (schematic representation), b) Substitute system.
k_A Damping constant of body shock absorber, c_A Spring stiffness of body spring,
k_T Damping constant of vibration absorber, c_T Spring stiffness of absorber spring,
m_T Absorber mass, m_A Body mass, z_A Vertical body motion, m_R Wheel mass,
z_R Vertical wheel movement, c_R Vertical tire spring stiffness,
k_R Damping constant of tire, h Vertical road excitation.



Wheel suspensions

Basic principles

Vehicle wheels and the vehicle body are connected via wheel suspensions. On the one hand, a wheel suspension has the function of guiding the relevant wheel in relation to the vehicle body in such a way that a movement that is essentially directed vertically relative to the vehicle body remains possible and, on the other hand, that the tire forces exerted in the wheel contact point in the horizontal plane and the torques generated by these forces can be transferred to the body. On the front axle and also on the rear axle of vehicles with rear-axle steering, additional steerability of the wheels is to be ensured.

Alongside the tires, the suspension and shock absorber system, the vehicle body mass, and the individual wheel masses, the wheel suspensions have a major influence on the driving characteristics of a vehicle, as they influence the parameters of the vehicle axle concerned that are relevant to driving dynamics. These are, for example:

- the track width,
- the toe-in or toe-out angle,
- the camber angle,
- the caster angle,
- the caster offset in wheel center,
- the spread angle,
- the spread offset,
- the kingpin offset,
- the disturbing force lever arm,
- the position of the rolling pole of the axle and thus the orientation of the roll axis,
- the location of the pitching pole,
- the braking and anti-squat control,
- the longitudinal and transversal springs.

For the definitions of the individual wheel-suspension and vehicle parameters, refer to the chapter “Chassis systems”, section “Characteristics”:

Kinematics and elastokinematics

During operation of the vehicle, the geometry and kinematics of a wheel suspension lead to changes in the characteristic parameters of the wheel suspension (for example, caster angle, location of the rolling pole) and wheel-position parameters (for example, camber angle and toe angle) of the corresponding wheel. This is the case, for example, as a result of a movement of the wheel within the framework of the vertical degree of freedom permitted by the wheel suspension (i.e., compression or rebound movement of the wheel) or with steerable wheel suspensions as a result of a steering movement. Figure 1 shows examples of the camber- and toe-angle changes of a wheel suspension due to kinematics during compression by the compression path Δz in comparison with the design position. As a rule, the kinematic wheel-position changes are shown with the help of diagrams of the compression and rebound path of the wheel. The so-called “wheel paths” for the camber- and toe-angle changes of the wheel suspension shown in Figures 1a and 1b are shown in Figure 1c.

Due to the fact that the kinematic changes to the wheel-suspension parameters and wheel-position parameters have a great influence on the driving characteristics of a vehicle, correct configuration and coordination of the steering and wheel lift kinematics are of great significance.

Alongside the changes to the wheel-position parameters due to kinematics during compression or rebound movements, the forces and torques affecting the wheel suspension (for example drive and braking forces, lateral and vertical forces in the wheel contact point) in conjunction with the elasticity of the suspension lead to further wheel-position changes. The elasticity of a wheel suspension results from the deformability of the individual wheel suspension components (for example

links) and the bearings used when forces and torques are applied. For reasons related to driving comfort and acoustics, the mounts used in modern wheel suspensions are usually elastic mounts (for example rubber mounts). Figure 2 shows an example of a wheel suspension with two rubber mounts fitted on the vehicle body side; their elasticity when longitudinal force occurs at the wheel contact point leads to an elastokinematic change in the toe angle of the wheel.

Alongside the kinematic wheel-position changes, the elastokinematic effects also influence the driving characteristics of a vehicle. When coordinating the kinematics and elastokinematics of a wheel suspension, this is why it is usually the aim that the kinematic and elastokinematic effects supplement one another when influenced by forces and springs.

For example, elastokinematic steering is used on a number of modern rear-axle suspensions to reduce load change reactions (for example by increasing the toe-in when braking force affects the rear wheel on the outside of the bend) [1]. It is possible to influence the elastokinematic properties of a wheel suspension by, for example, coordinating the individual mount elasticities or adjusting individual mounting points.

Figure 1: Kinematic wheel-position changes in the case of compression movement

a) Kinematic camber change (view from rear),
b) Kinematic toe-angle change (view from above),
c) Wheel path.
Δz Compression path from the design position,
$\Delta\gamma_{kin}$ Kinematic camber-angle change,
$\Delta\delta_{kin}$ Kinematic toe-angle change.

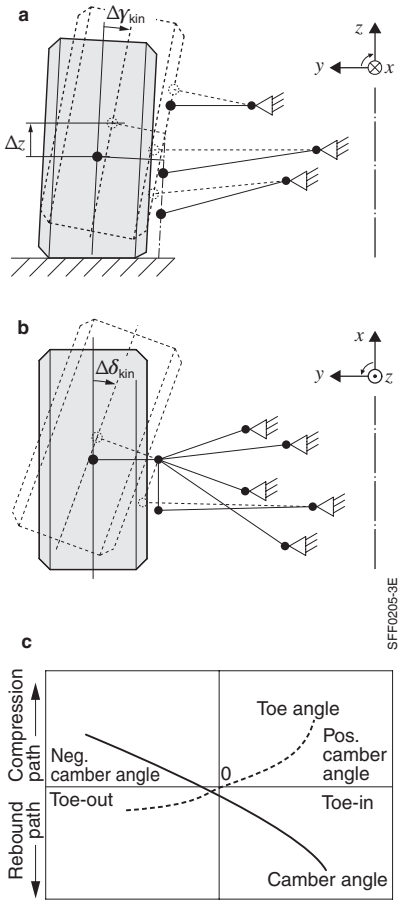
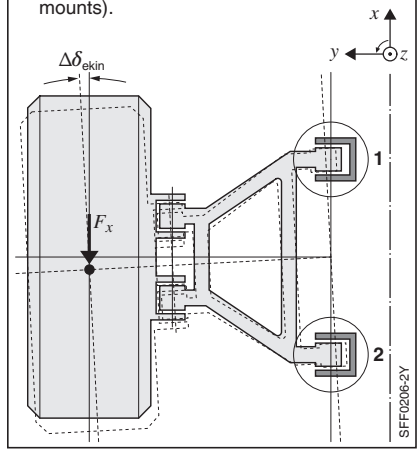


Figure 2: Elastokinematic toe-angle change due to effect of longitudinal force

F_x Longitudinal force (braking force),
$\Delta\delta_{ekin}$ Elastokinematic toe-angle change,
1 Elastic shift of pivot point 1 (rubber mounts),
2 Elastic shift of pivot point 2 (rubber mounts).



Basic categories of wheel suspensions

There are a large number of different wheel suspensions. They are classified primarily according to the type of suspension concept. An initial distinction is made between rigid axles (dependent wheel control), semi-rigid axles, and independent wheel suspensions (independent wheel control).

Rigid axles

In the case of a rigid axle, the wheels of an axle are firmly interconnected by a rigid axle body, which leads to mutual influences on the wheels. Rigid axles are used as both driven and non-driven rear axles on heavy vehicles (for example off-road vehicles, light utility vans, trucks). Occasionally, however, their sturdy construction and high ground clearance mean that steerable variants are also used as front axles (for example on off-road vehicles or off-road trucks).

Guidance of a rigid axle in relation to the vehicle body can be implemented in different ways. On vehicles with leaf springs, guidance is usually via the spring leaves (Figure 3a). There are also a large number of rigid axle concepts guided by links or coupling shafts (Figures 3b, 3c, and 3d). Where links and coupling shafts are used, statically undefined mounts are selected to make linking at the vehicle body easier and to reduce the required space [3]. Refer to [2] for detailed explanations of the individual axle variants.

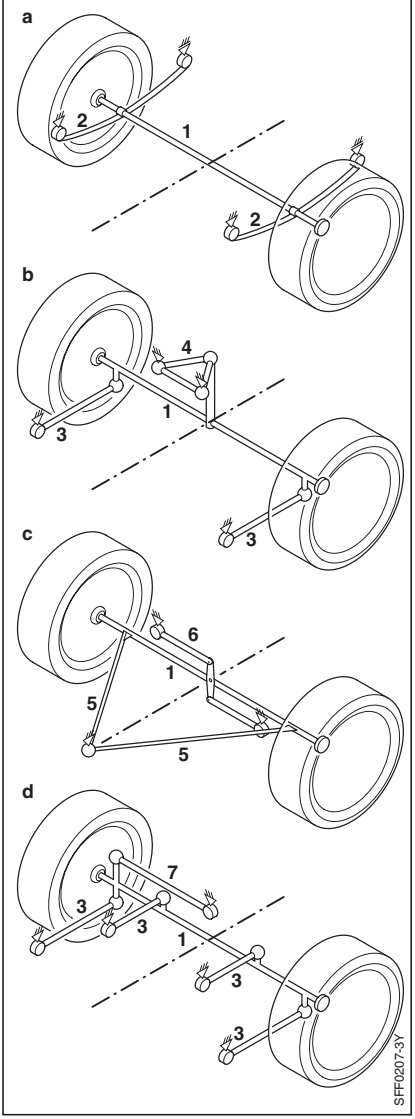
The major advantages of rigid axles are the simple and sturdy design, low costs, a high roll center, high maximum wheel lift, and high ground clearance.

However, rigid axles also have a number of disadvantages that are inherent in the design: the mutual wheel influence, high unsprung mass, high installation space requirement, as well as limited possibilities to coordinate the kinematic and elastokinematic factors.

Semi-rigid axles

Semi-rigid axles also involve a mechanical coupling of the wheels. In contrast to rigid axles, however, this coupling is not rigid. The elasticity of the coupling profile

Figure 3: Design examples for rigid axles
a) Longitudinal leaf-spring suspension,
b) Trailing and coupling-shaft links,
c) A-bracket with Watt linkage,
d) Trailing link with Panhard rod.
1 Rigid axle, 2 Leaf spring,
3 Trailing link, 4 Coupling-shaft link,
5 Coupling shaft, 6 Watt linkage,
7 Panhard rod.



that is used enables relative movements between the wheels. The coupling profile forms a cross-connection between two trailing links to which it is firmly connected. Longitudinal forces are absorbed via the trailing links. The bracing of lateral forces is supported by the stiffening effect of the coupling profile. In order to guarantee a relative movement between the two wheels of the axle, the coupling profile has to be designed as weak. Depending on the arrangement of the coupling profile, distinctions are made between torsional-link axles, twist-beam axles and semi-independent axles (Figure 4).

Due to their simple and low-cost design, semi-rigid axles are widely used as rear axles on vehicles with front-wheel drive.

The advantages of this axle concept include the low installation space requirement, the low unsprung masses, easy assembly and removal, the stabilizing effect of the coupling profile, the low track width and toe-angle changes, as well as the good anti-dive properties.

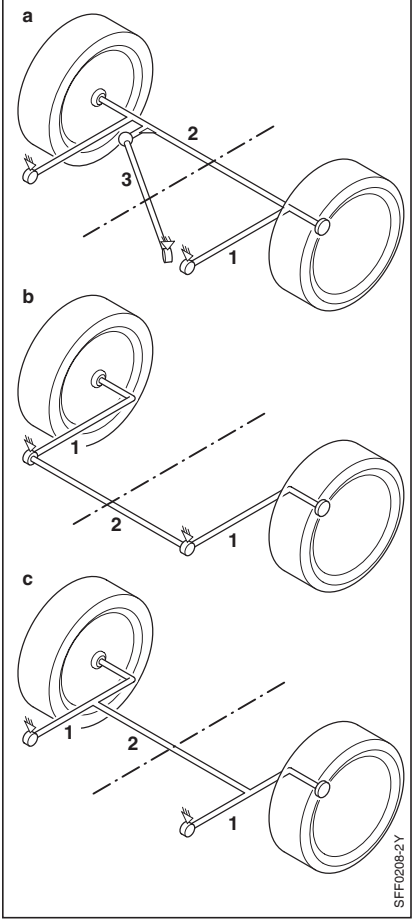
These advantages are offset by a number of disadvantages that are inherent in the principle: the mutual wheel influence, the low suitability for driven axles, the high tension peaks at the transition points between the trailing link and coupling profile, the increase in the tendency to oversteer in the event of influencing lateral forces (lateral force oversteer) due to link deformations, as well as limited kinematic and elastokinematic optimization potential.

Torsional-link axles
In the case of torsional-link axles (Figure 4a), the two wheel carriers are connected by means of a coupling profile arranged close to the wheel center. As a rule, the lateral guidance of the axle is supported by an additional guide element (for example a Panhard rod) [2]. There are great similarities to a rigid axle with regard to both the structure and properties.

Twist-beam axles
In contrast, twist-beam axles (Figure 4b) have kinematic properties similar to those of trailing-link wheel suspensions. The coupling profile is arranged at the height of the pivot points of the trailing link. The deployment and arrangement of the coupling profile greatly simplifies the mounting of the trailing links compared to a trailing-link independent wheel suspension.

Semi-independent axles
In comparison with twist beam axles, the coupling profile on a semi-independent

Figure 4: Design examples for semi-rigid axles
a) Torsional-link axle with Panhard rod,
b) Twist-beam axle,
c) Semi-independent axle.
1 Trailing link,
2 Coupling profile,
3 Panhard rod.



axle (Figure 4c) is not at the height of the link pivot points rather is offset towards the rear. This improves above all lateral force bracing in comparison with the twist-beam axle.

Independent wheel suspensions
Alongside semi-rigid axles – as rear axles on vehicles with front-wheel drive – most modern vehicles nowadays have independent wheel suspensions where each wheel is individually connected to the vehicle body according to the desired degrees of freedom of movement. A wheel is connected here with the help of a wheel carrier and a corresponding number of links.

The design (for example two-point links or A-arm links) and arrangement of the links (trailing links, transverse links or diagonal links) and the connecting mounts determine the kinematic and elastokinematic properties of the wheel suspension. The design of the individual links determines the number required to reduce the freedom of movement of a wheel to the desired number of degrees of freedom.

The number of links is frequently used as the classification of suspension types (for example, five-link independent wheel suspension). The resulting type of spatial movement (kinematics) of the wheel during compression and rebound movement is also frequently used for the classification of independent wheel suspensions [1], [3]. Depending on the type of movement of the wheel carrier, distinctions are made here between level, spherical, and spatial independent wheel suspensions [1], [3].

The proportion of independent wheel suspensions in modern vehicles is rising steadily. Compared to rigid and semi-rigid axles, independent wheel suspensions offer a number of advantages. For example, there is no mutual wheel influence, the kinematic and elastokinematic optimization potential is high, and the space requirement and unsprung mass are low in some cases.

However, independent wheel suspensions also have some disadvantages. In some cases, they lead to a complex design, the costs are high, the maximum wheel lift is low, and the configuration and coordination process is more complex in some cases.

There are large numbers of different designs of independent wheel suspensions. The basic principles of selected designs are to be explained briefly in the following section and their structures shown in diagrams. For detailed explanations of the individual designs and of other independent wheel suspensions and specific design examples, refer to [2].

Trailing-link independent wheel suspension
On a trailing-link independent wheel suspension, a wheel is connected to the vehicle body by means of a single link arranged in longitudinal direction (Figure 5a). The trailing link transfers both the longitudinal and lateral forces, which means that high mount forces occur and the mounts have to be designed accordingly.

The axis of rotation of the links runs parallel to the vehicle transverse axis. The advantages of this suspension form are usually low installation space requirement as well as low costs. The disadvantages are the limited kinematic optimization possibilities, the instantaneous center located at road height that causes high roll torque on cornering, as well as the high stresses placed on the links and their mounts.

Diagonal-link independent wheel suspension
In the same way as on the trailing-link wheel suspension, on the diagonal-link wheel suspension the wheel is also connected to the vehicle body by means of a single link. However, to achieve better bracing of the longitudinal and above all lateral forces, the link is arranged diagonally (Figure 5b) and there is more space between the mounting points. To achieve more favorable kinematic properties, on modern suspensions the axis of rotation of the link is arranged diagonally both in the projection to the vehicle lateral plane (roof angle) and in the projection to the road (V-shaped angle) [1].

Double-wishbone independent wheel suspension
A double-wishbone independent suspension is when a wheel is connected to the

vehicle body via two A-arm links. One link is arranged below the wheel center and the other is arranged above it (Figure 6a), which enables the suspension to brace all of the forces and torques that occur at the wheel. As a rule, the high articulation forces mean that the transverse links are not directly connected to the vehicle body structure but rather are secured to a “chassis subframe” that interconnects both wheel suspensions and thus relieves the load of inner forces on the vehicle body.

With adaptation of the mounts and design of the links, double-wishbone independent suspensions provide very high kinematic optimization potential [2]. Depending on the location of the axes of rotation of the links, level, spherical or spatial wheel suspension kinematics are achieved [3]. The disadvantages of double-wishbone independent suspensions are the higher costs as well as the greater installation space requirement.

Figure 5: Trailing- and diagonal-link independent wheel suspension
a) Trailing-link independent wheel suspension,
b) Diagonal-link independent wheel suspension.
1 Trailing link,
2 Diagonal link.

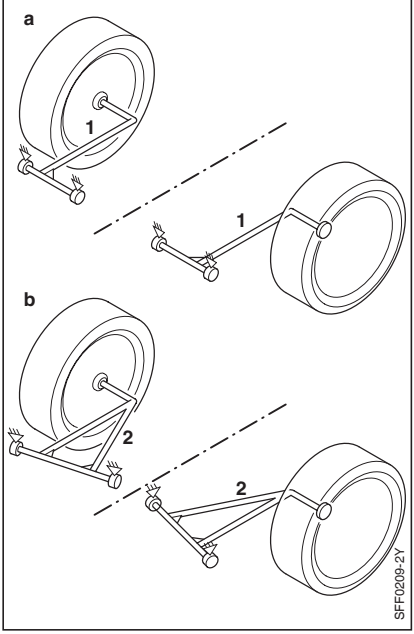
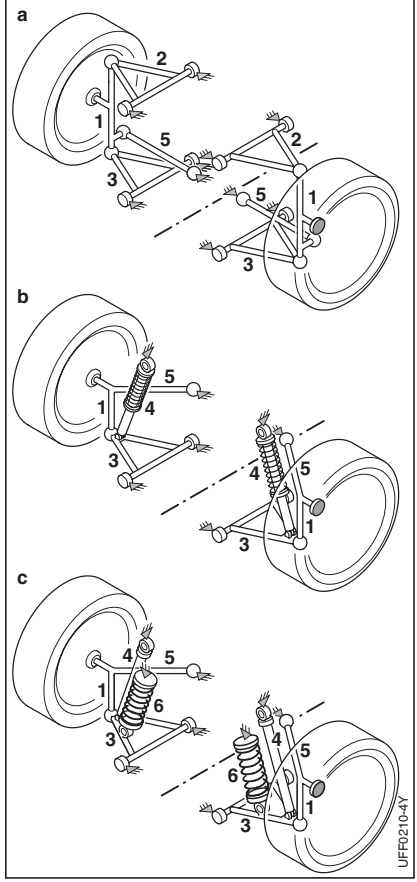


Figure 6: Double-wishbone and spring-strut independent wheel suspension
a) Double-wishbone independent wheel suspension,
b) Spring-strut independent wheel suspension,
c) Damper-strut independent wheel suspension.
1 Wheel carrier,
2 Upper A-arm link,
3 Lower A-arm link,
4 Spring damper strut,
5 Tie rod (steering).



Spring-strut independent wheel suspension

The kinematics of a spring strut wheel suspension correspond to those of a double-wishbone independent suspension in which the upper transverse link is replaced by a sliding guide (Figure 6b). This sliding guide corresponds to the spring strut on other suspensions (in the case of a combined spring-shock absorber unit) or the shock absorber strut (in the case of separate spring and shock absorber arrangement) where the housing is rigidly connected to the wheel carrier. With this design, the shock absorber rod also assumes wheel guidance tasks.

The lower link plane of a spring-strut suspension is usually formed by two two-point links (radius links) or an A-arm link. In the case of a suspension using the McPherson principle, originally the lower A-arm link was formed from a transverse link and a stabilizer. Nowadays, however, other spring-strut suspensions are also referred to as McPherson axles.

Advantages of the spring-strut wheel suspension are above all the low construction overhead and low amount of space required at the height of the wheel axes which can be used in particular in the case of passenger cars with front-wheel drive with laterally installed engine-transmission units. Other advantages are the cost-saving and weight-saving design, simple installation and assembly, as well as the high degree of integration. On specifying the kinematics, however, spring-strut suspensions offer slightly less leeway for configuration, among others of scrub radius and deflection-force lever arm, in comparison with double-wishbone independent suspensions.

Damper-strut independent wheel suspension

Kinematically similar to the McPherson axle, here too the supension-linkage function is assumed (Figure 6c) and requires a twin-tube shock damper to support the lateral forces. There are advantages in relation to the larger package requirement thanks to the omission of the spring-induced lateral forces on the damper (for improved response), reduced tendency to wear, and easier repair when changing the springs.

Multilink independent wheel suspension

Wheel suspensions with four or five single links are generally referred to as multilink axles. Multilink axles result, for example, from breaking down an A-arm link into two individual two-point links in the case of a double wishbone axle (Figure 7a). Breaking down three-point links and the use of independent two-point links usually result in greater leeway for configuration of the kinematic and elastokinematic properties of the axle. On the one hand, this increases the optimization of the axle with regard to comfort and driving safety requirements. On the other hand, the somewhat complex design increases the overhead involved in the configuration and coordination processes for the wheel suspension.

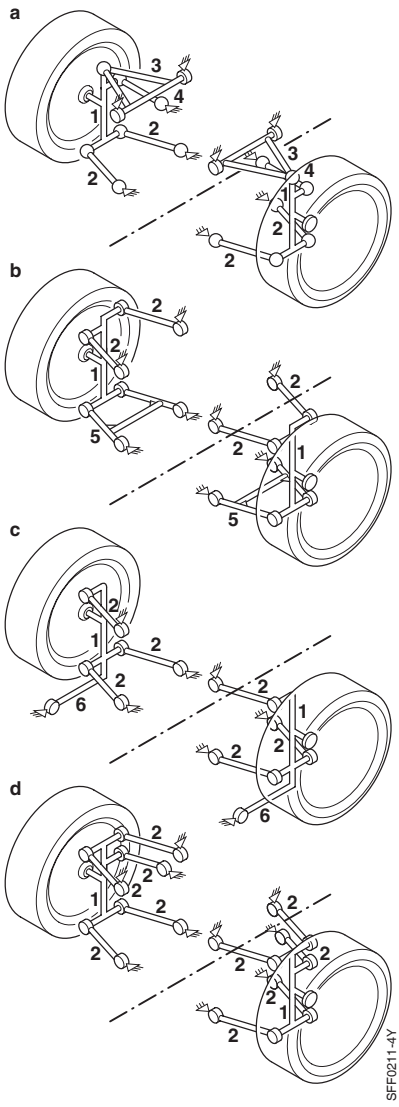
Trapezoidal-link independent wheel suspension

Trapezoidal-link wheel suspensions (Figure 7b) are a special form of the multilink independent wheel suspension that is mainly used as rear axles. The lower plane is formed by a trapezoidal link that has two connection points on the wheel side and an axis of rotation on the vehicle body side. This means the lower link sets a total of three degrees of freedom of the wheel. Another two degrees of freedom are eliminated by two two-point links arranged accordingly so that only the desired compression degree of freedom of the wheel remains.

Control-blade independent wheel suspension

Another design of multilink independent wheel suspensions is the control blade axle on which the wheel is guided by one trailing link and three transverse links (Figure 7c). The trailing link (control-blade link) is connected to the vehicle body in such a way that it can be rotated and is firmly connected to the wheel carrier. As a rule, it has elastic properties in order to enable kinematic track and camber changes. The lateral forces are braced via three two-point transverse links which are usually arranged in two planes (one above and one below the center of the wheel). The arrangement and orientation of the transverse links determines the kinematic properties of the wheel suspension.

Figure 7: Design examples for multilink independent wheel suspensions
a) Multilink independent wheel suspension,
b) Trapezoidal-link independent wheel suspension,
c) Control-blade independent wheel suspension,
d) Five-link independent wheel suspension.
1 Wheel carrier, 2 Two-point link,
3 A-arm link, 4 Tie rod (steering),
5 Trapezoidal link, 6 Control blade.



Five-link independent wheel suspension
A wheel suspension with completely detached links requires five individual two-point links to reduce the movement of a wheel to the desired vertical degree of freedom (Figure 7d). Five-link rear axles are generally referred to as multilink suspensions, whereby on the front axle this is referred to as a four-link axle plus tie rod [2].

References
[1] L. Eckstein: Vertikal- und Querdynamik von Kraftfahrzeugen. ika/fka 2010.
[2] B. Heissing, M. Ersoy (Editors): Fahrwerkhandbuch. Vieweg+Teubner Verlag, 2008.
[3] M. Matschinsky: Radführungen der Straßenfahrzeuge. Springer-Verlag, 2007.

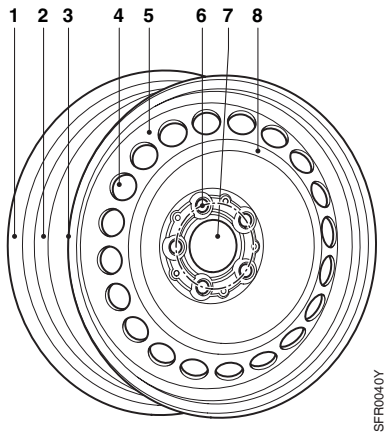
Wheels

Function and requirements

All of the vehicle-specific or axle-specific tasks are performed via the wheel, e.g. transfer of dynamic forces between the vehicle and road surface. These include taking up the vehicle load and the impact forces of the road surface, transferring the rotary motion of the axles to the tires, and taking up and transferring braking and acceleration forces as well as lateral forces when cornering. Wheel size is mainly determined by the space required by the braking system, the axle components, and the size of the tires used.

Wheels have primarily a technical function. But the increasingly booming light-alloy wheel market calls for visually attractive designs.

Figure 1: Structure of a disk wheel
1 Rim inner flange,
2 Rim base,
3 Rim outer flange,
4 Ventilation hole,
5 Wheel disk,
6 Pitch-circle diameter,
7 Center hole,
8 Dish.



Structure

The wheel is a load-bearing, rotating part between the tire and the axle. It usually consists of two main components – the rim and the wheel disk. These two components can be made from a single part, and can also be permanently or non-permanently attached to each other. A permanent connection of a rim with a wheel disk is called a disk wheel.

Figure 1 shows the basic structure of a steel wheel. Here the rim mounts the tire and the wheel disk connects the wheel to the axle.

In daily usage the terms rim and wheel are often interchanged. The term rim is often used when actually the complete wheel is meant. The “wheel” in general usage often also refers to the tire. However, as a technical term in automotive engineering “wheel” generally means the wheel without the tire.

Wheel disk
The wheel disk (wheel nave) is the part that connects the rim to the axle hub. In the case of a steel wheel, the wheel disk consists of a formed steel-sheet blank. This has holes for ventilating the brake system and is usually curved (dish, see Figure 1). The center of the wheel disk contains the center hole and the wheel-bolt or stud holes. The wheel is secured to the axle through these holes. The center hole is provided with a locating bore by means of which the wheel is radially centered on the axle. This bore determines together with the rim bead seat (for the bead ledge) the wheel’s true-running quality (via the radial run-out). The attachment face is, together with the rim flanges, responsible for lateral wheel run-out.

In the case of an aluminum wheel, the wheel disk can, due to its design, have different shapes such that the term wheel disk is no longer applicable.

Rim

The term rim strictly speaking describes only the radially outermost part of the wheel which holds the tire. The rim is therefore the fundamental connecting element between the wheel disk and the tire. In a tubeless tire it provides the air seal and is geometrically matched to the tire. The most commonly used shape of rim is divided into four zones (Figure 2):

– rim flange (inner and outer),
– hump (inner and outer),
– rim bead seat (inner and outer),
– rim base and drop center.

There are also different shapes for passenger cars (Figure 3).

Rim flange

The rim is limited on the inner and outer sides by a rim flange (rim inner flange and rim outer flange). It acts as the side stop for the tire bead (see Tires) and absorbs the forces resulting from the tire pressure and the axial tire load. The rim flange is specified in the guidelines of the ETRTO (European Tyre and Rim Technical Organisation) by, for example, J, K, JK or B. In this way, the geometry of the rim flange and the ratio to the drop center are dimensionally described. This is based on how the wheel is used.

The most common rim-flange shape for passenger cars is the J flange shape. The lower B flange can be found on smaller vehicles and in inflatable spare tire systems. K and JK flanges are rarely used any more; they used to be the domain of heavy, comfortable vehicles of the upper price segment.

Rim bead seat

The rim bead seat describes the contact zone of the tire with the rim. It centers the tire radially. In this zone the tire is given the correct position for radial and lateral true running. All the dynamic driving forces are transferred here. In the tubeless tires mainly used today in passenger cars the wheel/tire system is sealed at the rim bead seat.

Figure 2: Rim zones

1 Rim flange,
2 Rim bead seat,
3 Hump,
4 Rim base,
5 Drop center.

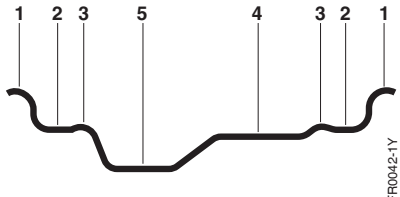
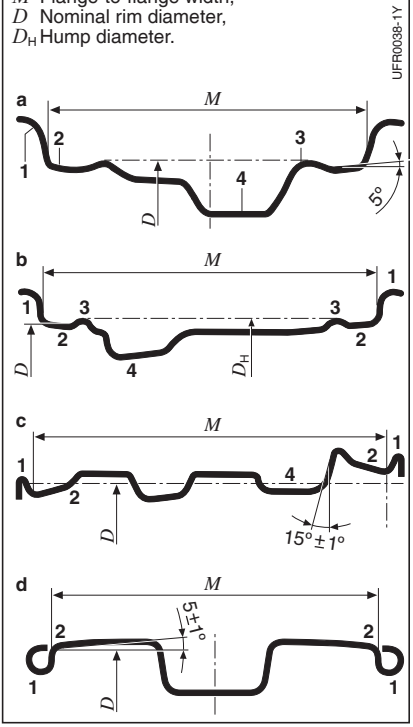


Figure 3: Rim systems for passenger cars

a) Passenger-car drop-center rim (standard rim),
b) EH2+ rim (extended hump),
c) PAX rim (Pneu Accrochage, X stands for Michelin radial-tire technology),
d) CTS rim (Conti Tire System), steel-wheel version.
1 Rim flange, 2 Tapered bead seat,
3 Hump, 4 Drop center.
M Flange-to-flange width,
D Nominal rim diameter,
D_H Hump diameter.



Rim base
The rim base connects the inner and outer rim bead seats. The drop-center rim is primarily used in passenger cars. Drop-center rims have a clearly defined shape with a drop-center rim base (drop center). When the tire is mounted on the wheel, the tire is initially positioned with one side of the tire bead in the drop center so that it can be pulled over the rim flange on the opposite side. The drop center is the required shaping of the rim base to mount the tire base (tire bead, tire inner ring) when fitting and removing the tire.

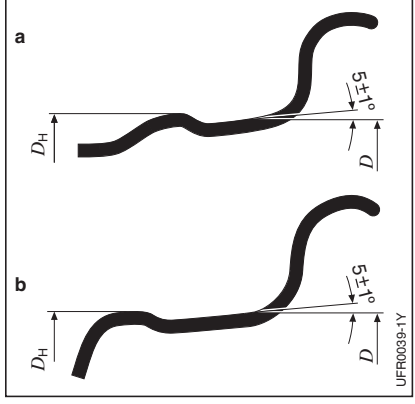
Hump
The hump is the all-round raised bead in the area of the rim bead seat (Figure 4). It is prescribed in many countries for tubeless tires. In the event of low tire pressure the hump is intended to prevent the tire from coming off the rim bead seat. The following hump shapes are customary:
– H: round hump on one side on the outer bead seat,
– H2: round hump on both sides,
– FH: flat hump on the outer bead seat,
– CH: combination hump,
  flat hump on the outer bead seat,
  round hump on the inner bead seat,
– EH2: extended hump on both sides.

Primarily H2 rims are used in passenger cars. H rims (formerly also called H1) are also used in commercial vehicles and older vehicles as well. The major differences are between the standard hump shape (H) and the flat hump (FH). Newly included in the standard is the extended hump (EH2) with slightly larger hump diameter, which is used in some cases, particularly with “run flat” tire systems.

Rim and wheel dimensions
Terms
The most important terms for the function and design of a wheel are (Figure 5):
– rim diameter (nominal diameter, dimension from rim bead seat to rim bead seat),
– rim circumference (measured value, determined with bead tape around the rim bead seat),
– flange-to-flange width (rim width, inside dimension between the rim flanges),
– center-hole diameter (centering diameter, as size of fit),
– rim offset ET (dimension in mm from the rim center to the attachment face of the disk wheel),
– pitch-circle diameter (diameter of the circle on which the center points of the bolt holes are situated),
– flange height (measured from the rim nominal diameter to the saddle point of the flange radius).

The rim offset determines the position of the wheel in the vehicle. The wheel is generally secured to the brake-disk chamber or to the brake drum. With a rim offset ET = 0 wheel attachment is exactly in the middle of the rim width. Changing the rim offset changes the position of the wheel and with it the track width of the vehicle. A smaller rim offset equates to a wider

Figure 4: Hump shapes
a) Hump,
b) Flat hump.
D Nominal rim diameter,
D_H Hump diameter.



track width, a larger rim offset equates to a narrower track width.

Rim size
The basic dimensions of a rim size using a commercial-vehicle disk wheel with 15° tapered bead seat rim by way of example is 22.5 × 8.25 inches. The first value gives the rim diameter in inches, the second value the flange-to-flange width in inches.

Rim designs
Depending on the intended purpose and the tire design, there are a range of different cross-sectional rim shapes available (Figure 6): Rims can be made from a single part or from multiple parts. According to EU standardization single-part rims must be identified with a "x" (e.g. 6J × 15H2) and multiple-part rims with a "-".

Drop-center rims for passenger cars and light utility vehicles
The rim base is recessed for tire mounting. The drop-center rim can be made

Figure 5: Rim and wheel dimensions
(Picture source: MAN Nutzfahrzeuge Group),
1 Bolt holes,
2 Rim center,
D Rim diameter,
ET Rim offset,
L Pitch-circle diameter,
M Flange-to-flange width,
N Center-hole diameter,
S Tire width.

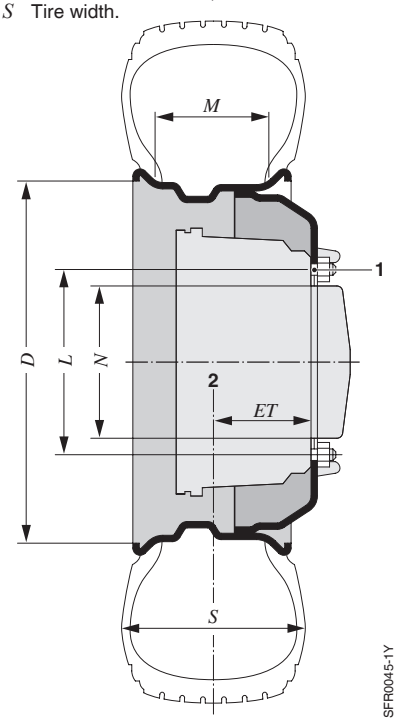
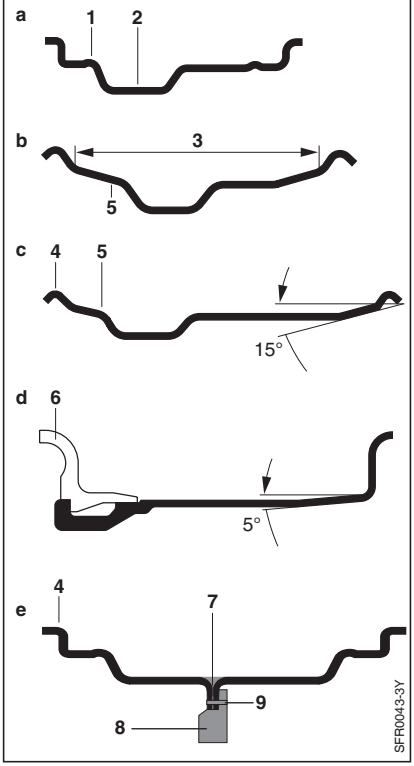


Figure 6: Rim design types
a) Drop-center rim,
b) 15° tapered bead seat rim,
c) Wide 15° tapered bead seat rim,
d) 5° tapered bead seat rim,
e) Two-part passenger-car rim.
1 Hump,
2 Rim base,
3 Flange-to-flange width,
4 Rim flange,
5 15° tapered bead seat,
6 5° tapered bead seat,
7 Sealant,
8 Wheel disk,
9 Bolt.



from one part (Figure 6a) or two parts (Figure 6e). The two-part rim is divided into a front half and a rear half which are bolted to each other in the drop center through an all-round pitch circle and to the wheel disk. A sealing ring or a sealing compound is used to provide the seal. The two-part variant traces its origins back to motor sport and offered the advantage of being able to replace one rim half in the event of damage. The tire is mounted in exactly the same way as on the single-part rim.

15° tapered bead seat rim
for commercial vehicles
The 15° tapered bead seat rim (Figures 6b and 6c) is made from a single part. The rim base is for tire-mounting purposes provided with a drop center to which the 15° tapered bead seats are connected. The 15° tapered bead seat rim with drop center is necessary to enable the benefits of the tubeless tire to be exploited on heavy commercial vehicles too.

5° tapered bead seat rim
The 5° tapered bead seat rim (Figure 6d), also called flat-base rim, is made from multiple parts. This is needed to mount the tires. The outer 5° tapered bead seat, which is non-permanently connected to the outer rim flange, can be removed. The outer 5° tapered bead seat is held on the rim base by an all-round sealing ring.

The tire is pushed onto the rim base when the 5° tapered bead seat is removed. The fact that the rim is made from multiple parts means that a tube is required. The two bead seats have a 5° taper.

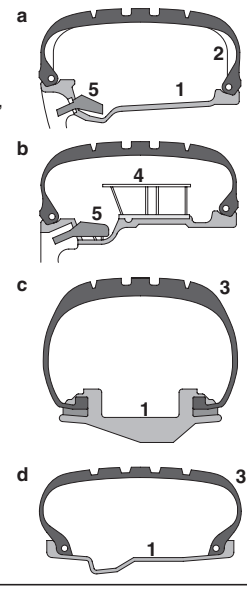
These rim systems have advantages when it comes to changing tires. However, they are heavier than 15° tapered bead seat rims and do not exhibit the good radial and lateral true-running properties of single-part rims.

Special trims
for special wheel/tire systems
PAX (Pneu Accrochage, X stands for Michelin radial tire technology) and CTS (Conti Tire System) are distinct rim geometries which can only be used with specially developed tires. This design is used primarily on armored vehicles. The complete system is designed to permit continued driving with a punctured tire without the tire coming off the rim. The two systems also prevent the tire from being destroyed by thermal load during driving when punctured. With a conventional rim base the sidewall folds up when the tire is punctured and friction occurs.

In the PAX system (Figures 3c and 7b) from Michelin an additional support ring is mounted on the rim base; the punctured tire rests on this support ring.

The CTS system from Continental does not have an additional support ring. The tire reaches round the rim base and can rest when punctured on the rim base (Figures 3d and 7c). Both run-flat systems are required by law to have a tire-pressure control system.

Figure 7: Special shapes for special wheel/tire systems
a) PAX system with reinforced run-flat tire,
b) PAX system with support ring,
c) CT system (cast-wheel variant),
d) Conventional wheel-tire system.
1 Rim,
2 Run-flat tire with reinforced sidewall,
3 Tire,
4 Support ring,
5 Valve with tire-pressure control.



Design criteria

Passenger-car wheels
Design criteria for passenger-car wheels include:
– High durability,
– Good support for brake cooling,
– Reliable wheel mounting,
– Low radial and lateral run-out,
– Small space requirement,
– Good corrosion protection,
– Low weight,
– Low costs,
– Problem-free tire fitting,
– Good tire seating,
– Good balance-weight seating (see Wheel with tire),
– Appealing design (for aluminum wheels),
– In part requirements to improve vehicle aerodynamics (c_d value).

Special requirements for commercial-vehicle wheels
Wheels for commercial vehicles are technologically highly sophisticated. Whereas racing-car wheels are geared towards maximum speed, commercial-vehicle wheels carry high tonnages at simultaneously – compared with excavators for example – high speed. In Europe, for example, long-distance commercial vehicles travel with a total weight of 40 tons at a speed of 80 km/h. It is essential when establishing the minimum tire size for a commercial-vehicle axle always to start out from the permissible axle load and the top speed dictated by the vehicle design.

The larger the load to be transported and the rougher the terrain to be negotiated, the more important it is to take into account the – sometimes competing – rim-design requirements. High load capacity is to be achieved by designing an appropriate wheel shape and using optimum materials. High fatigue strength is essential to increasing road safety. Low wheel weight optimizes payload and is required because the wheel, as an unsprung, rotating mass, influences the overall vehicle oscillatory system.

Designation for passenger-car wheels

A typical wheel designation for passenger-car wheels is, for example:
6 ½ J x 16 H2 ET30.

6 ½	Dimension of rim width in inches,
J	Rim-flange geometry,
×	Single-part rim base,
16	Dimension of rim diameter in inches,
H2	Rim hump at the inner and outer rim bead seats,
ET	Rim offset,
30	Dimension of rim offset in mm.

The designation and the associated dimensions with permitted tolerances have been prescribed as binding and standardized by worldwide recognized standards organizations such as ETRTO (European Tyre and Rim Technical Organization) or ISO (International Standards Organization) to harmonize rim and tire dimensions.

Materials for wheels

There are basically steel and light-alloy wheels. The material must always be specified in connection with the manufacturing technology. The following overview is intended to simplify systematization.

Classification of wheels

Steel wheel

A steel wheel consists of two parts, the rim and the wheel disk. Steel wheels are made from hot-rolled sheet steel in a forming process under rolling and bending conditions and joined by welding.

Light-alloy wheel

Light-alloy wheels are usually made of aluminum or magnesium alloys. They are produced using a variety of technologies. The aluminum wheel is produced as a cast wheel, forged wheel, sheet-metal wheel, or hybrid wheel. The magnesium wheel is only produced as a cast wheel.

The advantages of lightweight wheels are improved vibrational behavior, responsive suspension, reduced fuel consumption, and higher payloads. Light-alloy wheels are used in commercial-vehicle applications in particular in weight-sensitive transportation jobs. These jobs include the transportation of tanks and silos, where not exceeding the maximum transportation weight is of primary importance. In these situations the higher-cost light-alloy wheels usually pay for themselves within the first year of use.

Plastic wheel

Plastic wheels are made in an injection-molding process from mineral-fiber-reinforced polyamide and with metal inserts.

Materials

Sheet steel

In all the cheapest variant of passenger-car wheels is manufactured from hot-rolled and pickled sheet steel strip unrolled from the coil. The very good mechanical properties of this material provide for thin-walled wheel designed which are manufactured in highly automated, high-precision forming processes under bending conditions to finishing dimensions with close tolerances.

The continuous trend towards lightweight construction particularly since the CO_2 discussion has accelerated the use of high-strength, fine-grained structural steel. Its high tensile strength (600…750 N/mm^2) and very good formability and weldability enable lightweight, and cost-effective wheels to be manufactured efficiently.

Further potential for weight savings can be opened up by the use of “tailored blanks” for rim production. Here the sheet thickness of the starting material is adapted to the stresses in the wheel, in the course of which material strips of sheets of different thicknesses are joined by laser welding to create a blank.

Aluminum sheet

Aluminum sheet as an alternative to sheet steel can be more easily formed and is also – albeit thanks to more expensive methods (MIG welding) – more readily weldable. The manufacturing overhead is greater relative to steel wheels and the material costs are comparatively high, which prevents a wider range of applications. The use of high-strength sheet steels has greatly reduced the original weight advantage of aluminum sheet with the result that a cost/benefit analysis comes out in favor of steel wheels.

Light-metal alloys

Light-metal alloys are based in the main on aluminum alloys and in rare cases (e.g. in motor sport) on magnesium alloys. In the case of aluminum wheels, a distinction is made between cast and forge alloys depending on the manufacturing method used.

Cast alloys
Aluminum cast wheels are manufactured in low-pressure diecasting from aluminum alloys. The cast blank is formed in a steel mold, which is filled with liquid molten mass and cooled under controlled conditions to solidify. Aluminum alloys with a silicon content of between 7 and 11 % are used, depending on whether good castability or high strength is to be achieved.

Two alloys have proven successful. GK-AlSi11 is used for small wheels (up to 16 inches) with low wheel loads. The outstanding castability thanks to the high silicon content provides for highly efficient production with low reject rates due to casting defects. This alloy cannot be hardened by heat treatment. The wheels are therefore designed with greater wall thicknesses, which is reflected in slightly higher weights.

GK-AlSi7Mg is used for large wheels with a higher wheel load and for weight-optimized wheels. Adding 0.2…0.5 % magnesium to the aluminum alloy increases the strength of the cast wheel as a result of subsequent heat treatment (solution heat treatment and aging at warm temperature). This advantage is exploited to meet high load requirements during vehicle operation with a minimum of material usage.

To ensure that the high requirements with regard to strength, tightness and ductility placed on these safety-related vehicle components are met, only pure primary aluminum is used as the starting material. Contamination of the alloy with iron would cause acicular structures to form, which in turn would weaken the mechanical properties (elongation at fracture and tensile strength). Contamination with copper would reduce the chemical stability.

Forge alloy
Aluminum forged wheels are used in passenger cars when light, weight-optimized wheels are required and the target weight cannot be achieved with cast wheels. The increased hardness of the aluminum wrought alloy caused by the forging process (increase in mechanical strength due to plastic deformation) enables the wheel to be designed with a thinner wall, which means that less material needs to be used and thus the wheel is lighter.

The starting material takes the form of round continuously cast bars of AlSi1Mg which are sawn into precisely “portioned” disks. These are subjected to a three- to four-stage forging process to create the visible side (design side, disk, spider) and a flow-forming process to create the rim. As well as being hardened by plastic deformation the material is modified by a heat-treatment process.

Magnesium alloys
Whereas magnesium alloys have been unable to establish themselves in volume production – due to higher production costs in view of special safety precautions (risk of fire during cutting) – they are used in individual cases for special-purpose vehicles and racing cars.

Plastics
The use of plastic as a material for wheels is still in the developmental stage due in particular to insufficient high-temperature strength and difficult wheel mounting and manufacture. In particular the insufficient impact strength and thermal load capability as well as the incalculable long-term properties mean that plastic currently appears to make little sense as a material for a safety component such as the disk wheel in automobile construction.

Manufacturing processes

Steel wheels
When the text refers to passenger-car and commercial-vehicle steel wheels, exclusively sheet-steel wheels are what are meant. Other manufacturing processes such as casting and forging are not used for this material to manufacture wheels.

Passenger-car steel wheels consists of two parts: the wheel disk and the rim. They are welded to each other at the end of the production process. Both the wheel disk and the rim are manufactured in volume production on fully linked, highly automated production lines by means of forming processes under bending conditions. Smaller quantities can be produced more economically on individual presses.

Manufacturing the wheel disk
To manufacture the wheel disk the material required is unrolled directly from the coil, straightened flat and fed to a large transfer press with a force of pressure of approx. 40,000 kN. The press is equipped with a nine- to eleven-stage follow-on composite tool, in which the sheet blank is automatically routed with each press stroke from station to station.

A square blank with rounded corners is punched in the first machining step. Then in three to four stages the wheel disk and the center-hole area are formed by deep-drawing and stamping operations. In the next two or three stations the ventilation holes are punched with wedge driving tools and then stamped on the tool outlet side. The stamping replaces deburring of the pointed cutting edges and reduces the wheel's susceptibility to cracking under load.

Finally, the wheel designations are stamped on the reverse side and the calotte- or cone-shaped wheel-stud attachment faces are formed. In the last stage the wheel disk is calibrated to the finishing dimension for joining with the rim.

Manufacturing the rim
The rim material is likewise unrolled from the sheet coil, straightened and cut to length. The strips are stacked and forwarded to the automated rim production line. Here the sheet strip is shaped in a bending machine between three cylinder rolls into a ring and joined at the junction point in a butt-seam welding machine. The upsetting flash created in the process is then planed away on the inside and outside and the weld seam is smoothed by rollers. The unavoidable upsetting flash also on the side edges is then deburred and rounded.

The final rim contour is created on roller-burnishing machines by three consecutive forming steps, each with three profile rollers. The contour of the tool rollers is transferred to the rim here.

Where necessary, flow-forming is used to adapt the wall thickness and material distribution in the rim profile to the loads involved, thus saving more weight.

In the next operation the valve bore is produced on a rotary table. In the first station a flat attachment face is struck, from which the bore is punched in the second cycle. The cutting edges are then rounded on both sides by stamping.

Finally, the rim is drawn in a press onto a gage with the exact finishing dimension in order to create the tight concentricity and lateral-running tolerances (calibration).

Joining the wheel disk and rim
The process of joining the wheel disk to the rim creates the steel wheel. The two wheel parts – wheel disk and rim – are routed to a fully automated welding installation. On a small press they are aligned to each other and then joined to each other in a precisely specified position.

In the next station the assembly is connected by shielded arc welding with four to eight weld seams. These are deslagged and cleaned by rotating brushes adapted to the wheel. Then the wheel is calibrated to the finishing dimensions and made available for surface treatment.

The manufacturing process for steel wheels is so precise that the blank wheel is able to move on to the painting plant without any mechanical reworking. There, as a general principle, it runs through cathodic dip coating and, if stipulated, is given an additional visually appealing finishing coat.

Sheet-steel wheels are characterized in particular by their robustness and, thanks to their low manufacturing costs, are offered in the entry-level equipment specs of vehicle models.

Aluminum wheels

Aluminum-sheet wheels

For the most part, the basic manufacturing process is identical to that of the sheet-steel wheel. In view of the lower strength, the wall thickness must be greater compared with the sheet-steel wheel. Despite engineers having full mastery of the technology, aluminum-sheet wheels have not gained acceptance because steel wheels as a whole are the more economical wheel variant and the design options are very limited.

Aluminum forged wheels

As the name suggests, a forged wheel is created by the hot forming of an aluminum round blank between two forming tools. The forming process takes place in two stages – the forging of the wheel front and rear sides between two defined tools and the flow-forming of the rim contour.

The starting material for aluminum forged wheels takes the form of 6 m long continuously cast bars of AlSi1Mg with a diameter of 200…300 mm, depending on the planned wheel size. After an ultrasonic test the cavity-free bar sections are sawn off to a predefined length. The saw support – a cylinder approx. 250 mm in diameter and 150 mm high – is fed to the automated forging line. This consists of a heating furnace and up to four consecutive forging presses with forces of pressure of 8,000…40,000 kN. Parts are handled between the presses by robots. The result of this first forming process is blank with the finished design of the wheel disk, a punched-out center hole and a ring-shaped material reservoir positioned around the circumference with the material scheduled for the rim.

In flow-forming operation involving three rollers the ring-shaped disk is split open (hence the term “split wheel”) and rolled out on a bell-shaped tool to make the rim.

Before being machined, the forged blank is subjected to heat treatment to improve the mechanical properties. The complete wheel contour is turned on two consecutive lathes (i.e. machined to finishing dimension). Then the bolt holes and the valve hole are drilled and milled on a machining center. The concluding polishing process gives the wheel its reflecting luster.

This high-precision machining ensures that every wheel runs absolutely true. There is no radial or lateral run-out.

If small batch sizes do not justify the manufacture of a forging tool, a special form of forged-wheel production is used. This involves the manufacture of a forged blank in the shape of a thick-walled cylinder whose base is a disk representing the rotation contour of the design and of the inner side of the wheel. With high machining overhead, the wheel design is usually 100% milled and the rim base first flow-formed and then turned.

Aluminum cast wheels

The most common method used is low-pressure mold casting. In the casting machine the aluminum molten mass tempered under controlled conditions is located in a crucible underneath the mold. The mold and crucible are connected by means of a feed tube. Once the mold is closed, the pressure in the crucible is increased to approx. 1 bar, causing the molten mass in the feed tube to rise and fill the mold.

Fusion heat is drawn off during the solidification process through precisely defined cooling channels in the mold. The specific cooling and heat removal during the solidification process and the casting parameters in their entirety (pressure, temperature and time) are decisive factors for the casting quality.

The production process is completely automated from the casting stage. The cast blanks removed by robot arms pass via linked conveyors through the following machining stages until they are automatically stacked and packed as wheels in the dispatch area:
– casting,
– removal of riser bore,
– X-ray test,
– heat treatment,
– machining,
– brushing and deburring,
– leak test,
– painting,
– dispatch.

In the X-ray test all the blanks are tested for casting defects according to specifications stipulated by the customer. The defective parts with casting defects that are not visible from the outside, e.g. porosity or shrink holes (cavities, material breaks) and inclusions or contaminants, are separated out and returned to the melting furnace.

The automated run is interrupted only before the painting stage so as to form production batches which are painted the same color.

Flow-forming process
Where necessary, it is possible to use a modified, slightly more sophisticated process to manufacture weight-optimized cast wheels, so-called “flow-forming wheels”. This produces a weight saving of around 0.9 kg for a 19″ wheel. For the flow-forming process, the cast blank is made in a similar way to that used in forging. A ring is provided around the design surface instead of the formed rim contour as a material deposit for the rim. This ring is processed in a specially designed machining cell as follows:
– preturning for rolling,
– heating,
– rolling out of the rim (flow-forming).

The blank formed in this way is returned to the “normal” manufacturing process prior to heat treatment.

Another way of manufacturing weight-optimized wheels is to insert lost cores in zones of the wheel subject to less stress. Thus, aluminum is replaced by cavities, for example in the spoke, but also much more rarely in the hump. Accordingly there are hollow-spoke wheels and hollow-hump wheels.

Squeeze-cast process
The squeeze cast process attempts to exploit the advantages of diecasting for aluminum wheels. An exactly portioned amount of molten aluminum is pressed under high pressure into a diecasting mold under exactly defined casting parameters. The great advantage lies in the high solidification speed with positive effects on the material structure. Other advantages are the significantly lower machining overhead – and thus less material usage – and the relatively high output and longer mold service life. This casting process is used in individual cases, but it requires special, relatively complex casting machines and molds. This process has yet to gain acceptance.

Wheel design variations

Single- or multiple-part design variation

Sheet-steel and aluminum-sheet wheels consist of two parts. In this design wheel disk and rim are welded to each other. In the case of forged light-alloy wheels and cast wheels, the one-piece version dominates.

Multiple-part versions, even those which are made of different materials (e.g. magnesium wheel disk and aluminum rim), are available mainly for tuning applications and for sports vehicles. Multiple-part wheels trace their origin back to motor sport. Mechanics exploited the advantage of being able to replace the damaged parts. In tuning applications, for example, the possibility of standardized rim rings and disks is used to create a large number of different wheel dimensions. For the most part, however, multiple parts no longer have a technical background and are now only used for visual reasons. Multiple-part wheels are subdivided into two- and three-part wheels.

Wheel spider

The wheel spider on cast wheels refers to the area of the spokes, which is represented in a steel wheel as the wheel disk and is for the most part provided with openings. Ventilation holes, slits and openings in the wheel disk or the wheel spider serve on the one hand to reduce weight, and on the other hand to ventilate the brake system and to enhance the visual design of the wheel. Set against this is the present-day requirement also to optimize the effects of the wheel with regard to the entire aerodynamics of the vehicle. Depending on the dynamic behavior of the body, a positive effect can be achieved here by keeping the area of the openings small and fashioning the geometry of the spokes as flat as possible. This measure has a more negative impact on the weight of the wheel such that it is now more common to find plastic parts on an aluminum wheel which are intended to aerodynamically optimize the wheel.

Rim variants

Rims used for passenger cars, vans and light utility vehicles are almost always drop-center rims with H2 double humps (rarely with FH or FH2 flat humps), tapered bead seats, and J section rim flange. Less common on smaller vehicles is the lower flange shape B, which today is used primarily in compact spare wheels. The higher flange shapes, JK and K, are rare on modern vehicles, and then only on heavy vehicles.

Lightweight-construction technologies

Hollow-spoke technology with sand casting cores or ceramic cores remaining (lost) in the wheel also provides good possibilities to reduce weight, but it requires a suitable design and special manufacturing equipment. This process is also associated with higher costs.

More widely used today is the flow-forming process for cast wheels, in which the rim base is only partly precast and then rolled out by machine to the corresponding rim width. The compressed material provides for thinner wall thicknesses with reduced weight in the rim base.

"Structural wheels" are used, among other things, as spare wheels or as road wheels with plastic wheel covers. Unrestricted by design conditions, the aim here is to use the minimum quantity of material possible to guarantee operating and functional safety, as well as streamline production costs for these wheels.

Wheel mounting
The design of the wheel and the mounting elements must meet safety requirements in all vehicle operating conditions. The wheel forces resulting from motive force, brakes, wheel load and wheel location must be supported by the overall mounting system (wheel bolt, wheel hub, brake-disk chamber, wheel-bolt holes, possibly coatings of parts) without impairing fatigue limits or the function of the wheel and axle components. Careful coordination of the friction parameters and geometry at the wheel bolts or the wheel nuts and contact zone of the wheel (bolt head to wheel-bolt hole) is essential when specifying the tightening torques in engineering design and in practice.

The geometric configuration of the wheel mounts in pitch diameter, number and dimensioning of the mounting elements are subject to the needs and requirements of each vehicle manufacturer. The wheel on a passenger car is secured to the axle hub by three to five wheel bolts or wheel nuts inserted through the mounting holes. Off-road vehicles and light utility vehicles often have six wheel bolts or wheel nuts. Commercial vehicles as a rule have ten wheel nuts, but this number can sometimes be even higher (e.g. tractors and excavators). The design of the contact surface of the nuts varies, depending on the vehicle manufacturer (e.g. calotte, cone, flat head). The longitudinal bolt forces that are decisive for the durability of the bolted connection must be reached and adhered to, both when new and used, in all dynamic operating states.

The high degree of true-running is achieved by means of a central wheel mount at the wheel hub with a precise alignment shoulder.

At present, wheels mounted with a central nut and interlocking driving pins are used almost exclusively on racing cars.

Wheel trims
Wheel trims (wheel caps) are mainly used for visual reasons on steel wheels and are affixed to the wheels using elastic retaining-spring elements which are easily detachable. But today aluminum cast wheels also feature wheel trims, which are often intended to improve aerodynamics. The design of aluminum wheels is for the most part kept simple and the weight kept low. Bolted solutions are also used in rare cases. The material used most often for wheel caps is heat-resistant plastic, e.g. polyamide 6. However, in some cases, aluminum and stainless pressed steel are also used.

Special wheel/tire systems
TRX rim
More recent rim developments, which have produced in limited series, are the TR rim (in metric dimensions). They were developed by MICHELIN for use with matching TRX tires and provide more room for the brakes.

Rims made by DUNLOP with a Denloc groove also require special tires; at low tire pressure and also in the case of pressure loss, the system is supposed to prevent the tire from coming off the rim, enhancing safety and mobility.

The TD system (TRX-Denloc) brings together both wheel/tire systems. As opposed to common practice, the two above designs are have rim and tire matched to each other and are unable to combine rim and tire with other tire versions, or only to a very limited extent.

CTS and PAX systems
The CTS/CWS and PAX systems were able to dispense with the need for a spare wheel. The original idea behind using these two systems to save on the spare wheel failed to catch on in the market and today these are used primarily on armored vehicles (see Rim designs).

Compact spare wheels
For space-saving reasons a compact spare wheel (mini spare) is often used as the spare wheel. This can be stowed in combination with a collapsible spare tire in an even smaller space (e.g. in roadsters, convertibles). All compact spare wheel systems are equipped with a specially designed tire where the driving properties are only suitable for emergency operation and limited top speed (approx. 80 km/h). Its benefits are the subject of debate, but it is becoming increasingly popular compared with a full-size spare wheel.

In many countries it is no longer required by law to carry a spare wheel. Instead, vehicles are equipped with a puncture kit (Tire-Fit) to repair any tire damage. The puncture kit consists of an electrically driven compressor and a sealant, which is pumped into the tire through the valve.

Stress and testing of wheels

The extremely varied and complex stress conditions in the wheel as a component in conjunction with a wide variety of operating conditions in the vehicle require specific endurance tests in order to be able to confirm the durability of a wheel with acceptable overhead. In general, the dynamic tests are run in test laboratories on standardized testing units, whereby a simulation of road operation that is close to reality is simulated and a good correlation of the test results to pure road operation is achieved. Country-specific legal requirements make special tests necessary, e.g. in the case of light-alloy wheels the simulation of a side curb impact (impact test).

Testing of sheet-steel wheels
The critical zones on a sheet-steel wheel are in particular the zones around the weld seams, mounting boreholes, dish (curvature of the wheel disk, see Figure 1), and ventilation holes. The operating conditions in each case, for example straight-ahead driving and cornering, generate different damage patterns in the area of the welding seam on the drop center well and in the wheel disk. Tests of the material quality and of the welded joints as well as surface tests are backed up by the endurance tests and indicate the need to optimize wheel manufacturing.

Testing of light-alloy wheels
Light-alloy wheels run through a similar testing process whereby, unlike sheet-steel wheels, the more varied influencing parameters of material, manufacturing, and design mean that the test requirements are at a significantly higher level. This ensures that fluctuations in material and manufacturing are unable to lead to premature failure. The maximum stresses occur mainly on the back of the wheel in the supporting structure of the ribs and spokes, in rare cases on the visible side.

The material quality and processing have a great influence on the durability of aluminum cast wheels. Inadequate physical values such as elasticity (during expansion) and tensile strength can be caused by poor heat application during casting or during heat treatment. This leads to porosity and shrink holes and deficient structural formation. The burrs that occur in high-stress zones during machine-cutting represent preliminary damage similar to notches and are often the starting point for incipient cracks. Careful machine deburring of these zones or specific constructive countermeasures, e.g. generously molded radii, are essential.

Testing of wheel with tire
Concentricity and lateral running
To assess the concentricity (true running) quality of a wheel on a vehicle, it is necessary to assess the wheel with the tire fitted, i.e. as a wheel with tire. During the manufacture of a wheel for concentricity the hub centering is in proportion with the two areas for the inner and outer tire seats. Likewise the contact face on the wheel hub and the inner areas of the rim flanges are responsible for the lateral running of the wheel. As a result of production these areas are endowed with tolerances (0.3 mm is usually given for concentricity and lateral running for passenger-car wheels), with which the tolerances of the tire now overlap. This can positively or negatively influence the concentricity of the wheel with tire. “Matching” is used to facilitate optimum concentricity of a wheel with tire. In this process the wheel and the tire are positioned during fitting in relation to each other in such a way that the “concentricity high point” of the wheel matches up with the “low point” of the tire.

On the wheel the high point is determined from the concentricity measurements of the two tire-seat areas. For each area an individual high point is obtained with different angular positions on the circumference of the wheel. These two values produce through vector addition a common value with a resulting angular position. This position is marked on the wheel with a colored dot or an adhesive dot.

On the tire the low point corresponds to the position where it reaches the lowest force variation while rolling. It is also marked with a colored dot. From a technical viewpoint the tire can also be compared with a spring that exhibits a radial stiffness. Based on production the tire can never be manufactured to such precision that it exhibits a uniform stiffness over its entire circumference. Wheels with tires with poor concentricity make themselves felt on the vehicle not only through a radial movement of the body (i.e. in the z direction). In the direction of travel a minimal variable force from acceleration and braking is also experienced with each wheel rotation.

Good wheel centering is extremely important for commercial vehicles that travel at higher speeds, but also when the wheels in question are large and heavy. On commercial vehicles that travel at higher speeds in particular, the lowest possible radial and lateral run-out on both rim bead seats and flanges is essential to ensuring smooth running. This increases safety and fuel economy.

Imbalance
Just as important as concentricity and lateral running to a smooth-rolling wheel with tire is compensation of the differently distributed masses on the wheel and tire. To this end it is necessary to minimize the influences of the masses on the rotating

wheel with balance weights by balancing. Normally passenger-car wheels are dynamically balanced due to the rim width, i.e. measurements are taken on two planes (inner and outer tire seats) and the required compensation mass is determined. This is then applied with balance weights at the point indicated by the balancing machine. Adhesive, clip-on or drive-on balance weights are used for this purpose (Figure 8). The ideal position for the balance weights on the wheel for dynamic balancing is the maximum distance to the rim center at as large a diameter as possible.

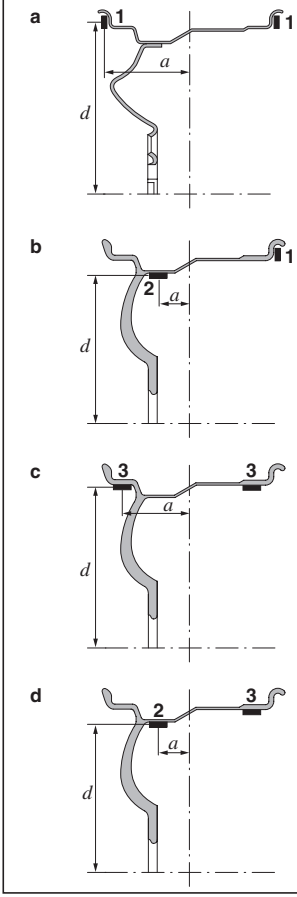
An imbalance, referred to as “residual imbalance”, of 5 g per balance plane depending on the vehicle type and suspension is not discernible in the majority of vehicles. For each balance plane a balance weight should be attached at one position only. If a relatively high balance mass (over 80 g) is required in a balance plane, it is recommended to turn the tire on the wheel and repeat the balancing process. The lower the balance-weight mass on the wheel, the lower the residual-imbalance potential also.

Narrow wheels for two-wheeled vehicles are balanced on one plane only (with a balance weight in the rim center). This method is called static balancing.

Wheels for commercial vehicles and compact spare wheels with a restricted top speed are not balanced.

Figure 8: Balance-weight positions

a) Clip-on or drive-on weight on the inside and outside (visible),
b) Clip-on or drive-on weight inside in combination with an adhesive weight under the drop center (concealed),
b) Clip-on or drive-on weight inside in combination with an adhesive weight under the tire seat (concealed),
d) Two clip-on drive-on weights (concealed).
1 Balance weight with retaining spring clipped onto the rim flange,
2 Balance weight glued to the inside of the wheel drop center,
3 Balance weight glued to the inside of the rim bead seat.
a Distance of balance weight to wheel center,
D Distance of balance weight to axis of rotation.



Tires

Function and requirements

The tire is the only component of a vehicle that comes into contact with the road. It thus assumes a key driving-dynamics position. Downstream driving-dynamics control systems such as the Antilock Braking System, the Traction Control System and the Electronic Stability Program are only ever as effective as the tire allows within the framework of its instantaneous power-transmission potential. When it comes to the tire, ultimately the crucial factor is the active safety of the vehicle.

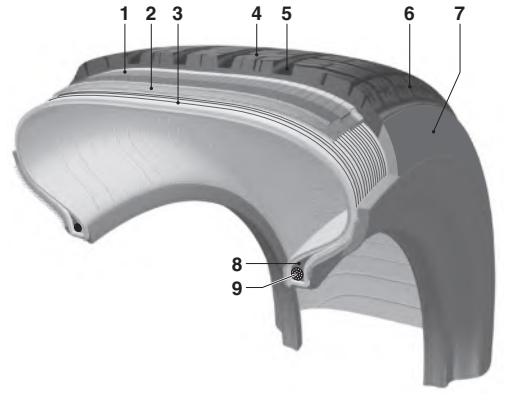
Tires perform a variety of functions in everyday driving applications: They cushion, damp, steer, brake, accelerate, and simultaneously transmit forces in all three dimensions – at high and low temperatures, in the wet, on dry roads, on snow, mud and ice, on asphalt, concrete and pebble stones. They are meant to roll straight, permit precise steering, absorb road irregularities, bring the vehicle safely to a stop, and be quiet and comfortable. They are also meant to last, retain their characteristics with increasing age and decreasing tread depth, and produce as little rolling resistance as possible. Furthermore, an inflated tire performs supporting, vibration-damping and comfort-giving functions, and is therefore an active and fully integrated suspension element. The requirements resulting from these functions to be met by tires can be summarized as follows:

– high-speed resistance,
– durability,
– abrasion resistance (mileage),
– low rolling resistance,
– good properties in the wet (aquaplaning, wet braking, wet handling),
– good tire comfort, quiet rolling noise,
– good running characteristics in the limit range,
– resistance to aging,
– precise steering properties (handling),
– short braking distances,
– simple installation and assembly,
– true running and uniformity,
– economy,
– resistance to damage,
– resistance to chemicals.

The fundamental and visible material of a tire is rubber, an elastic to viscous material to which a tire owes the majority of its typical properties that are so important to the vehicle.

Figure 1: Tire construction

1 Nylon binding,
2 Steel-belt assembly,
3 Radial textile cord plies (casing),
4 Tire tread (tread rib),
5 Tread (tread groove),
6 Tire shoulder,
7 Sidewall,
8 Bead apex,
9 Bead with bead core (steel core, numerous thin steel cables twisted to other).



Tire construction

Design and components
Tires are a complex construction of different, mutually influencing raw materials, components and chemicals. A standard passenger-car tire consists of up to 25 different components and up to 12 different rubber compounds.

Today, only the tubeless, radial-ply steel-belted tire constructed in two stages satisfies the stringent demands imposed by the automotive industry and consumers.

Ingredients
The ingredients of a radial tire are:
– natural and synthetic rubber (approx. 40 %),
– fillers such as e.g. soot, silica, silane, carbon, and chalk (approx. 30 %),
– strength members such as e.g. steel, aramide, polyester, rayon, and nylon (approx. 15 %),
– softeners, e.g. oils and resins (approx. 6 %),
– vulcanization accelerators, e.g. sulfur, zinc oxide, stearin (approx. 6 %),
– anti-aging agents, e.g. UV and ozone blockers (approx. 2 %).

Since 2010 the toxicologically serious softeners and paraffined oils have been subject to particularly stringent limits values in the EU. For this reason manufacturers are increasingly using uncritical natural oils (e.g. sunflower oil).

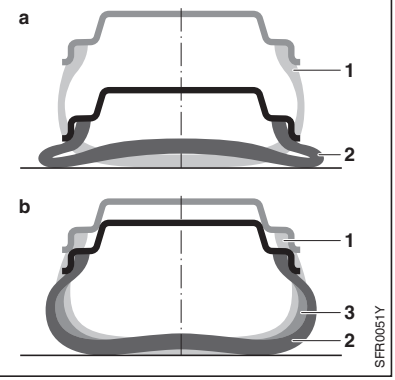
Casing
The casing is stretched over a thin inner liner of airtight butyl rubber (Figure 1). Around 1,400 rubberized cords of rayon, nylon or polyester are combined in one or more casing plies to form the decisive strength member, the elastic “shell” of the tire. The cords run radially, i.e. at right angles to the tire plane from bead to bead – hence the designation radial tire. Cross-ply tires, in which casing cords are placed diagonally to the tire plane, in practical terms no longer play a role in modern-day applications.

Bead
The bead performs the important function of ensuring that the tire is securely and tightly seated on the rim. Driving and braking torques are transmitted via this crucial connecting point from the rim to the tire tread and thus to the road surface. Seated in the bead core is a cable of numerous steel wires, each of which can bear a load of up to 1,800 kg [1].

Sidewall
A thin and highly flexible rubber flank forms the sidewall and thus the flexible zone of the tire. The sidewall (tire flank) is however also the area of the tire that is most sensitive to damage.

Trouble-proof run-flat tires on the other hand have much thicker sidewalls than conventional designs (Figure 2). In the event of a blow-out the rim does not slump onto the tire casing and thus cannot damage it. In addition, fully deflated run-flat tires ensure a certain degree of steerability and directional stability for a further 80 km at speeds of up to 80 km/h.

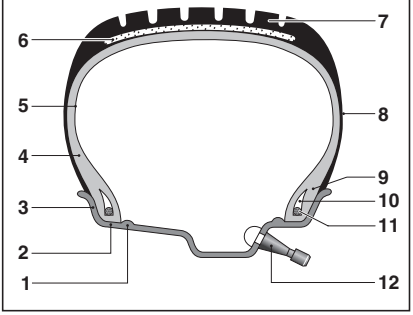
Figure 2: Comparison of run-flat tire with standard tire
a) Standard tire,
b) Run-flat tire.
1 Tire with normal inflation pressure,
2 Tire without inflation pressure,
3 Reinforced sidewall.



Cambering to blank
The finished, cylindrical combination of casing, inner liner, bead and sidewalls is pushed over a drum, the outside diameter of which corresponds to the inside diameter of this preliminary tire stage and to that of the subsequent tire. On this drum the cylindrical combination is cambered into the "real" tire shape (inflated and fixed) and then built further.

Because the casing cords run radially, i.e. transversely, to the rolling direction, the casing on its own would not be able sufficiently to transmit lateral forces when cornering and peripheral forces when accelerating and braking. It therefore needs support. This job is performed by the steel-belt assembly placed on top. Two or more plies of twisted and brass- and rubber-coated steel wires (steel cord) run not in the peripheral direction, but alternately at acute angles of between 16° and 30° to each other. High-speed-resistant tires are additionally stabilized by a nylon or aramide binding which suppresses a peripheral increase caused by centrifugal force. The tread surrounds the casing.

Figure 3: Design of rim with tire
1 Hump, 2 Rim bead seat,
3 Rim flange,
4 Casing (cord carcass),
5 Airtight rubber liner,
6 Steel belt,
7 Tread,
8 Sidewall,
9 Bead (with bead foot, bead core and bead apex),
10 Bead apex,
11 Bead core with steel core,
12 Valve.



From blank to finished tire
The tires in this penultimate stage is called a blank and is now placed in a heating press. Inside this press is an exchangeable recess – an exactly shaped negative mold of the later finished tire. In this heating mold the tire blank is "baked" under steam pressure (approx. 15 bar) and heat (to 180 °C) for up to 30 minutes and acquires its final, typical appearance. The tread rubber creeps on heating exactly and without cavities into the tire negative mold of the heating press, thus creating the tread pattern and the sidewall markings. As a result of sulfur added in a previous process the hitherto plastic rubber vulcanizes into elastic rubber and acquires its desired operating characteristics.

The tread pattern ensures low rolling resistance, water expulsion, good grip, and high mileage.

A finished tire of the standard size for the medium-size car class 205/55 R 16 91 H weighs around 8.5 kg. A commercial-vehicle tire of the standard size 385/65 R22.5 weighs around 75 kg.

Tire with wheel
Together with the rim, the tire valve and the balance weights the tire forms the car's operational wheel (Figure 3). The rubber-elastic tire is inflated with compressed air, at which point it is able to absorb and transmit forces. Tire inflation pressure is usually 2…3.5 bar for a passenger car and 5…9 bar for a commercial vehicle. It is not the tire itself but rather the inflation air that supports the weight of the vehicle.

Differences between commercial-vehicle and passenger-car tires
Commercial-vehicle tires and generally similar in design to passenger-car tires, but are bigger, wider and heavier. Tire inflation pressure at 5…9 bar is much higher than approx. 2…3.5 bar for passenger-car tires.

The primary development objective is, as with passenger-car tires, to have all the parameters in proper proportion and above all the mileage. Commercial-vehicle tires therefore have comparatively hard, low-wear treads which are also regroovable and retreadable. Retreading of a bald tire identified as such is possible if the tire casing is undamaged.

Although the rolling resistance of a truck tire is lower than that of a passenger-car tire, its influence on the truck fuel consumption is greater due to the higher vehicle weight and the number of axles. In addition to high load capacity (up to 3…4 t per tire), other important tire characteristics include good straight-running stability, good cornering stability and traction. The trend in truck tires is towards increasingly smaller tire dimensions. This increases the useful load height and thus the transport volume.

Tire inflation pressure

Motor-vehicle manufacturers specify two values for tire inflation pressure for every vehicle – the part-load air pressure for a partly laden vehicle and the full-load air pressure for a fully laden vehicle or high driving speeds. The values are based first and foremost on the vehicle weight, on its top speed, on the tire design, and on the tire size. As a rule, the front and rear axles are subject to different desired pressures. The tire inflation pressure may only be measured and adjusted on a tire that is cold, i.e. that has not been heated up by vehicle operation.

A correct tire inflation pressure is important for
– optimum tire contact patch and optimum ground contact,
– shortest possible braking distance,
– optimum wet grip,
– balanced cornering stability,
– low rolling noises,
– low rolling resistance,
– low flexing work and heat production.

An excessively low tire inflation pressure will result in each case in the opposite end of the above parameters and in
– reduced service life,
– increased and sometimes uneven abrasion,
– progressive structural damage,
– danger of sudden tire blow-out,
– increased risk of accident,
– increased fuel consumption.

An excessively high tire inflation pressure (much less critical compared with low pressure)
– results in impaired tire comfort,
– reduces the ground contact patch (tire “stands”) and thus diminishes cornering-stability and braking-force potential,
– causes increased central abrasion,
– but only reduces the rolling resistance slightly.

Tire tread

A tire is provided over its entire circumference with geometrically shaped tread grooves, ribs and channels as well as additional notches forming gripping edges (sipes) (Figure 4).

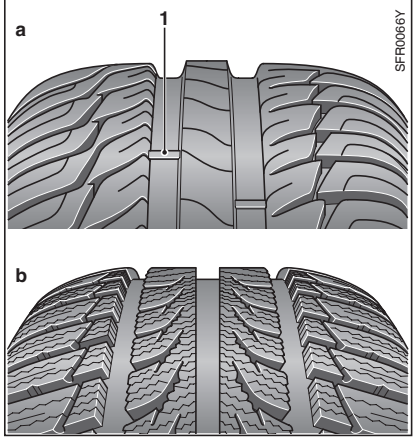
The most important function of the integrated tread (not cut, but heated) is to adequately absorb and disperse water on the road surface (also snow and mud in the case of winter tires), since wet and even surface moisture alone have a negative effect on grip characteristics. The braking distance is therefore dependent not only on the effects of interlocking and adhesion produced by the two friction partners tread rubber and road surface (see Tire grip). The braking distance on wet roads increases as tire wear increases.

Minimum tread depth

Summer tires

The minimum tread depth prescribed by law (EU Directive 89/459 of 1989 [2]) is established in most European countries at 1.6 mm (for passenger cars).

Figure 4: Tire tread
a) Typical tread of a summer tire,
b) Typical tread of a winter tire.
1 Wear marker.



Winter tires

The minimum tread depth for winter tires varies greatly from country to country. In Austria, for example, winter tires for passenger cars are required to have a minimum tread depth of 4.0 mm.

Wear-detection aid

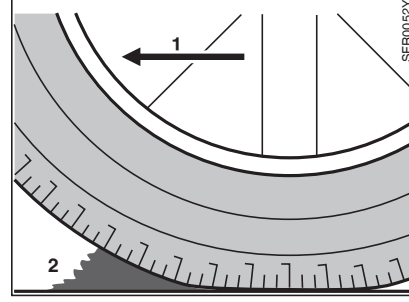
The main tread-groove base features several rubber bumps exactly 1.6 mm high spread across the entire tread for the purpose of checking the minimum tread depth prescribed by law of 1.6 mm. If such a wear marker (TWI, tread wear indicator) on the tire makes contact with the road surface, the tire may no longer be used in road traffic.

Truck tires bearing the wording "Regroovable" on their sidewalls may be regrooved by the regrooving depth approved by the tire manufacturer (depending on the tire version 2…4 mm), ideally when the remaining tread depth is still 2 to 4 mm. Essentially, passenger-car tires are not permitted to be regrooved.

Aquaplaning

At higher speeds or when there is an enclosed layer of water on the road the tread the tread is no longer able to absorb sufficient water in itself and disperse it to the sides and to the rear. A wedge of water forces its way between the tires and the road, the tires lose contact with the road, and the vehicle loses its controllability – aquaplaning occurs (Figure 5).

Figure 5: Aquaplaning
1 Driving direction,
2 Splash area.

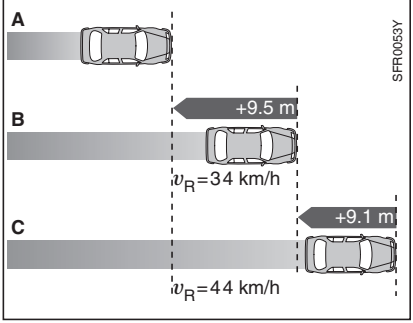


The tire loses contact at precisely the stage when the pressure of the wedge-shaped splash water in front of the tire exceeds the pressure of the tire on the road. This pressure is squared as the driving speed increases. Because the critical pressure at which the tire aquaplanes is approximately equal to the internal tire pressure, passenger-car tires with an inflation pressure of approx. 2.3 bar aquaplane at a much lower speed than truck tires with 8 bar. Driving with a lower inflation pressure than prescribed decreases the already low speed from which aquaplaning occurs, once again significantly in passenger-car tires.

Tire contour, tread design and tread depth can defer the speed from which the risk of aquaplaning arises. Narrow tires, due to the higher pressure on the road (surface pressure, weight per contact patch), essentially aquaplane at higher speeds than wide tires, and they also heave to channel a much smaller volume of water. Drainage channels and rounded contact surfaces on wide tires reduce their risk of aquaplaning to an acceptable level. To compare: A tire with a nominal width of 220 mm must at 80 km/h and a rainwater height of 3 mm disperse roughly 15 liters per second so as not to aquaplane. This figure is 10 liters for a 140 mm “narrow” tire.

Figure 6: Braking distance from 80 km/h to a step on a wet road with new tires and with bald tires [3]

A Braking distance with 8 mm tread depth: 42.3 m.
B Braking distance with 3 mm tread depth: 51.8 m.
C Braking distance with 1.6 mm tread depth: 60.9 m.
v_R Residual speed.



Wet braking

Figure 6 shows the further point at which the tread depth is of crucial importance to road safety: The braking distance on a road wet with rain is roughly 50% longer with an almost bald tire (tread depth 1.6 mm) than with a new tire of identical size (tread depth 8 mm).

The residual speed v_R denotes the speed of the worse braking vehicle at the moment when the vehicle with better tires comes to a stop. It is calculated as

$$v_R = \sqrt{v_0^2 \cdot \left(1 - \frac{s_1}{s_2}\right)} \text{ in m/s,}$$

where
v_0 Driving speed at start of braking in m/s,
s_1 Braking distance with vehicle 1 (with better tires) in m,
s_2 Braking distance with vehicle 2 (with bald tires) in m.

The residual speed is a measure of the theoretical accident severity to be expected in the event of a collision.

These calculated differences however can in practice only be effected by highly experienced drivers with extremely fast reactions. Many car drivers are not accustomed to ABS full braking; the total braking distance is lengthened significantly. Potentially shorter-braking tires only get a look-in if the vehicle is fitted with a brake assistant.

Force transmission

The four tire contact patches (footprint) are the direct and sole interface between the road surface and the vehicle.

Slip angle and slip
Only the tires rolling under an angle to the wheel rolling plane (slip angle α, Figure 7) and in the process deforming and constantly more or less slipping (see slip) transmit simultaneously and within physical limits the forces requested by the driver through steering, braking and accelerating. Conversely: A tire that does not roll or slip at an angle does not transmit any forces.

The tire transmits ever higher forces as the slip angle and slip increase. This relationship however is not linear. The effect is reversed after a relevant maximum value is reached (Figure 8). For passenger-car tires this reversal point for the slip angle is roughly 4…7°, corresponding to a pronounced steering-wheel angle. The reversal point for slip is 10…15 % (on snow up to 30 %). Excessive slip is a consequence of excessive acceleration or excessively heavy braking. If the steering angle or brake pressure is increased further, the wheels lock if the vehicle is not fitted with an antilock braking system. The slip is then –100 %.

Longitudinal and lateral forces
If a force F_x in the peripheral direction and a side force F_y occur simultaneously (e.g. when braking while cornering), the resulting transmitted horizontal force

$$F_h = \sqrt{F_x^2 + F_y^2}$$

cannot exceed the value $\mu_h F_z$. This situation can be explained by reference to the Kamm (friction) circle (Figure 9). The radius of the Kamm circle is equal to the maximum horizontal force $\mu_h F_z$ that can be transmitted via the tire. The maximum side force F_y is therefore smaller is at the same time a force F_x occurs in the peripheral direction. With the forces F_x and F_y marked in Figure 9 the wheel is exactly at the limit of the maximum horizontal force that can be transmitted.

Figure 7: Slip angle
1 Wheel rolling plane,
2 Tangent to driving direction,
3 Driving direction.
α Slip angle.

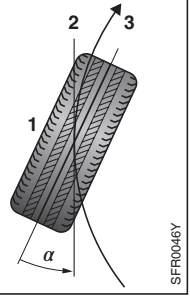


Figure 8: Coefficient of friction vs. slip

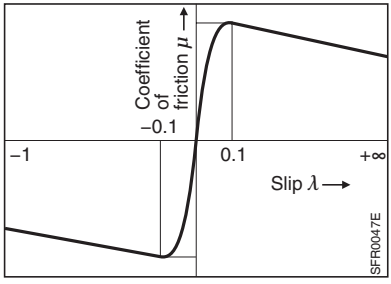
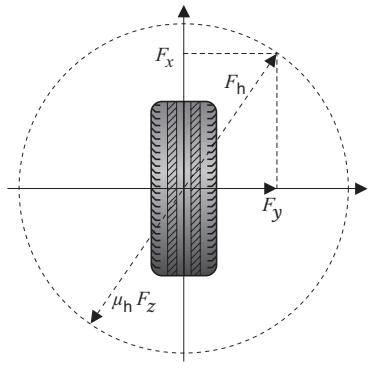


Figure 9: Kamm circle
F_x Peripheral force,
F_y Side force,
F_z Normal force,
μ_h Coefficient of static friction.



Tire grip

Generation of grip

Tires must transmit all the dynamic forces to only four roughly postcard-sized areas. The grip required for this purpose between the tire contact patches and the road surface is generated by several simultaneously occurring phenomena. These are essentially positive locking (also referred to in this context as the interlocking effect) and adhesion by molecular forces of attraction.

When a vehicle driving past at a constant speed is considered from the outside, as the tire rolls the continuously changing ground contact patch remains apparently fixed in relation to the vehicle (Figure 10) – while each individual rubber block of the tire tread runs into this forcibly flattening contact patch, deforms, and is "ejected" again at the other end. Relative movements and thus slip are generated in this contact patch: Each individual rubber block more or less slips during the dwell time in the contact patch.

Figure 10: Flattening in the contact patch between tire and road surface

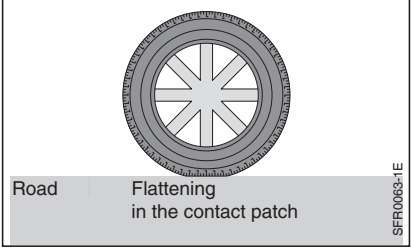
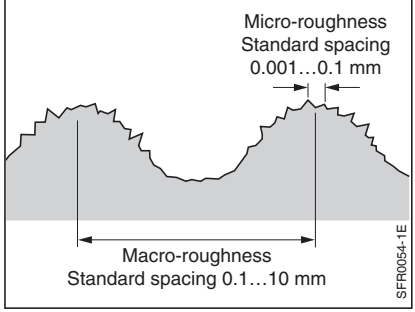


Figure 11: Micro- and macro-roughness of a dry road surface



Viscoelasticity

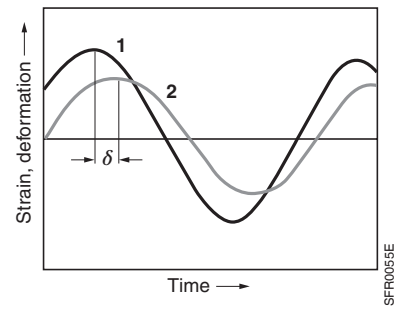
Viscoelasticity describes the time-, temperature- and frequency-dependent elasticity and the viscosity of polymer and elastomer substances (e.g. of plastics, rubber). Internal damping, molecular interlocking and creeping processes prevent the two extreme states "fully elastic" (e.g. like an elastic spring) and "highly viscous" (like a solid body). Deformation and the force that causes it as well as the mechanical strain pass off at different times.

Interlocking effect

The interlocking effect is created by the direct and intensive contact of the tire with the road, depending on the micro- and macro-roughness of the road surface (Figure 11). In the passage of the contact patch the tread block under consideration comes up against a bump in the asphalt, is upset, and slips off again on the other side of the bump at accelerated speed. Only when it produces slip in the process can it build up in the tangential direction a counterforce contrary to the rolling direction which counteracts the sliding and thus permits the transmission of steering, drive or braking forces.

Figure 12: Load application and load removal of a viscoelastic substance

1 Strain, force per unit of area,
2 Deformation, elongation or upsetting relative to output variable.
δ Phase lag.



On account of its viscoelastic properties a rubber block does not revert immediately to its original shape after deformation; the strain follows on from the deformation that causes it (Figure 12). This effect typical of rubber of hysteresis results, on account of the cyclic deformation of the viscoelastic rubber, in a loss of energy in the form of non-utilizable heat and thus in a contribution to friction (hysteresis friction). The component of the friction force parallel to the road surface thus facilitates the transmission of drive or braking forces.

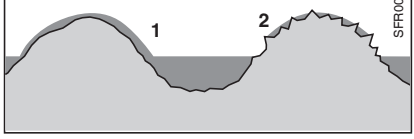
The principle of positive-locking grip also functions on damp and wet micro- and macro-rough road surfaces, but with limited effectiveness. Figure 13 shows that microfine asphalt points can penetrate the layer of moisture and thus the interlocking effect is maintained. However, an enclosed layer of water forms over the more rounded local bumps.

Essential to the interlocking effect is the presence of micro- and macroscopically small road irregularities. On a completely smooth surface (the coefficient of friction μ approaches zero) there would be no interlocking effect at all.

The interlocking effect and inner-molecular friction cause the tire to heat up. The resulting loss of energy is jointly responsible for the tire's rolling resistance, which makes up around 20…25 % of a vehicle's fuel consumption.

Figure 13: Micro- and macro-roughness on a damp or wet road surface

1 Macro-roughness channels and stores the water, but cannot penetrate the water layer.
2 Micro-roughness generates local pressure peaks and can thus penetrate the residual layer.



Adhesion

Molecular adhesion is created by the interaction and intensive contact between tire and dry road. The creation and subsequent breaking up of adhesive connections at the contact points result in a contribution to the coefficient of friction (adhesion friction). On a wet road molecular adhesion fails while the interlocking effect remains effective.

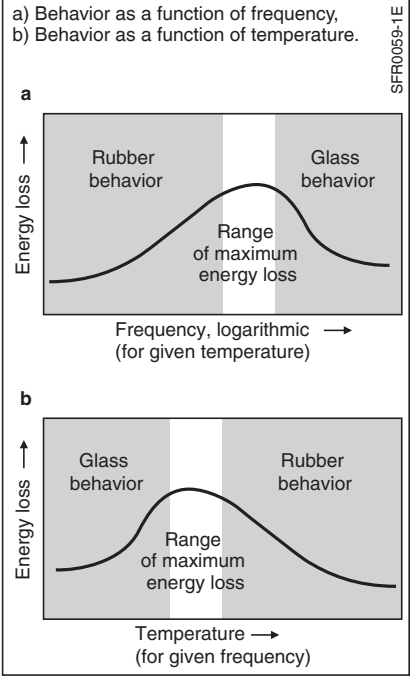
The frequency range of molecular adhesion, excited by the micro-rough road surface during drastic braking and extreme cornering, encompasses the spectrum of 10^6…10^9 Hz.

Load frequency and temperature

Two further important influencing factors on the quality of grip must be mentioned: If the tire when rolling is only excited with a low frequency (exposure frequency), the rubber behaves elastically (low energy loss, Figure 14a). In this low frequency range the rolling resistance is very low, the tire is relatively cold, and the grip is weaker. If, on the other hand, the frequency increases due to the excitations caused by micro- and macro-roughness, a viscoelastic behavior is manifested – the ideal range for tire grip (energy loss has its maximum). If the frequency continues to increase, viscosity (flowability) and energy loss decrease again, the is barely able to be deformed and hardens (glass behavior).

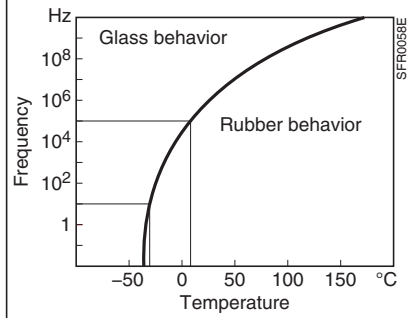
In parallel, rubber part exhibits pronounced heat-dependence behavior: In the glass-temperature range – when for example winter ambient temperatures cause the rubber compound to harden and become brittle (hence the analogy to glass) – the coefficient of friction of the tire rubber and with it the energy loss decrease markedly as a result of the rubber's molecular immobility (Figure 14b). Conversely, higher forces can be transmitted when the tire is moved in its optimum operating-temperature range.

Figure 14: Energy loss in the tire as a function of frequency and temperature
a) Behavior as a function of frequency,
b) Behavior as a function of temperature.



For rubber, therefore, an inversely proportional dependence between temperature increase and exposure frequency can be ascertained. Thus the glass temperature of an elastomer rises from –20 °C at only 10 Hz to +10 °C at 10^5 Hz (Figure 15). Compound developers are able to design rubber compounds with a glass temperature of between –60 °C and 0 °C at a frequency of 10 Hz.

Figure 15: Rubber behavior as a function of frequency and temperature



Special case: Rubber friction
Coulomb's classic law of friction ($F_R = \mu m g$) does not apply to vulcanized tire rubber. Its coefficient of friction μ (also called friction value or friction factor) is not constant, but
– increases with decreasing surface pressure (force per contact patch, given in N/mm^2),
– decreases with increasing surface pressure,
– depends on the sliding velocity,
– depends on the temperature and exposure frequency.

The phenomenon of increased grip with decreasing surface pressure explains the importance of very wide (and thus low-surface-pressure) tires in motor sport.

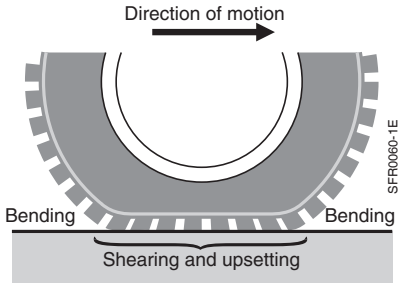
Rolling resistance

Definition of terminology
Tractive resistances inhibit the forward movement of the vehicle and must be overcome by motive means. In addition to the aerodynamic drag, the frictional resistances in moving engine, transmission and chassis/suspension components, the climbing resistance, and inertia forces, the rolling resistance of tires is classed as one of the main tractive resistances. It contributes to approx. 20 % of the fuel consumption on expressways/interstates, approx. 25 % on orbital roads and approx. 30 % on urban and ordinary roads [4].

Rolling-resistance decreases thus result directly in consumption and emission reductions. Rolling resistance (RR) corresponds to the energy loss per unit of distance and is given like every force in N (newtons). The dimensionless rolling-resistance coefficient c_{RR} denotes the ratio of rolling-resistance force to vehicle weight.

Example: Assuming a rolling-resistance force F_{RR} = 120 N and a vehicle weight G = 10,000 N ($G = mg$; vehicle mass m in kg, gravitational acceleration $g \approx 9.81$ m/s^2) c_{RR} amounts to 0.012 = 1,2 %. Standard values for c_{RR} for passenger-car tires on asphalt amount to 0.006 to 0.012, the (lower) value for truck tires to 0.004 to 0.008.

Figure 16: Shearing, upsetting and flexion in the contact lead-in



The "unit" kg/t (kilogram per ton) is occasionally used. In the example above the result 0.012 is 12 kg/t. This means in the case of a wheel load of 1 t = 1,000 kg that the rolling-resistance force F_{RR} assumes a value of 120 N.

Rolling-resistance-optimized tires have additional designations such as "Eco", "Green" or "Energy". Tire designers could indeed immediately and quite significantly reduce the rolling resistance by choosing rubber grades with low hysteresis and thus a low energy loss, but, as already explained, this would reduce the grip values to unacceptable levels. In other words, lower fuel consumption comes at the expense of less grip.

Generation of rolling resistance
With each wheel rotation the tire is deformed during the forcible flattening in the contact patch by flexion, upsetting and shearing of the rubber blocks and of the proportional tire casing (Figure 16). The fabric plies of the tire rub against each other (flexion), during which the tire performs flexing work. This produces a viscoelastically caused energy loss in the form of non-utilizable heat. This heat loss makes up 90 % of the rolling resistance.

Narrow tires and higher tire inflation pressure do increase the rolling resistance because contact patch and flexing work are reduced. However, designers are subject to very tight limits in terms of their scope of action in that the catalog of requirements with regard to handling performance, grip level and comfort are in direct conflict. Nevertheless, the technical specifications of the tire industry for future tire generations already feature tire dimensions such as 115/65 R 15 for subcompact-size cars and 205/50 R 21 for medium-size cars. Further drastic rolling-resistance reductions are not achievable with the requirements still standard today with regard to vehicle size, weight, maximum speed, sportiness, and comfort.

Even increasing the tire inflation pressure by 1 bar above the recommended value will only deliver a rolling-resistance reduction of 15 %. In a vehicle with an assumed fuel consumption of 10 l/100 km this would only produce a saving of 1.6 %. But because drivers still neglect to check tire pressure, the rolling-resistance reductions achieved in the most recent development cycles are not always implemented in real driving conditions. In addition, there is a direct conflict of aims between rolling-resistance optimization and the wet grip essential to road safety.

Conflict of aims between rolling resistance and grip

Frequency ranges

The deformations of the tread and rubber blocks caused by micro- and macro-roughness in the contact area between road surface and tire surface which generate the grip potential due to viscoelasticity occur in the very high frequency range of $10^3 \ldots 10^{10}$ Hz. The energy loss caused by hysteresis is high here, resulting in the creation of high grip values.

However, the frequency spectrum which is important to rolling resistance is much lower at 1…100 Hz – precisely the range in which the inner tire structure is excited with each wheel rotation. At a driving speed of 100 km/h a passenger-car tire is deformed roughly 15 times per second, equating to a load frequency of 15 Hz. Tire designers nonetheless refer here to low-frequency excitation of the tire structure, particularly the casing.

Optimization incompatibility

These greatly differing frequency ranges explain the basic optimization incompatibility of simultaneously high values for grip and low rolling resistance. In conventional tires using industrial soot as the primary filler (up until the mid-1990s) a rubber compound with high hysteresis in the high-frequency grip range automatically results in a high energy loss in the tire components subjected to low-frequency load and thus in high rolling resistance.

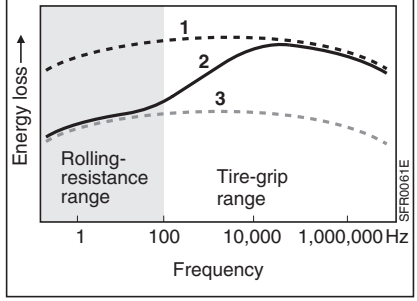
Silica to resolve the conflict of aims

The solution in the late 1990s involved the introduction of silica (trade name for refined silicate) as a gray-powdered filler, which increasingly replaced the standard industry soot used up that point. Together with auxiliary binding materials, or silanes, the conflict of interests between rolling resistance, grip and abrasion resistance can be elevated to a high level of compromise.

Silica-based rubber compounds exhibit low energy losses in the low-frequency range relevant to rolling resistance, but high energy losses in the high-frequency rubber-grip range (Figure 17). The curve for energy absorption has this rise steeply and thus advance in frequencies of $10^2 \ldots 10^4$ Hz. This results in tires which have low rolling resistance but nevertheless very good grip.

Figure 17: Frequency dependence of energy loss

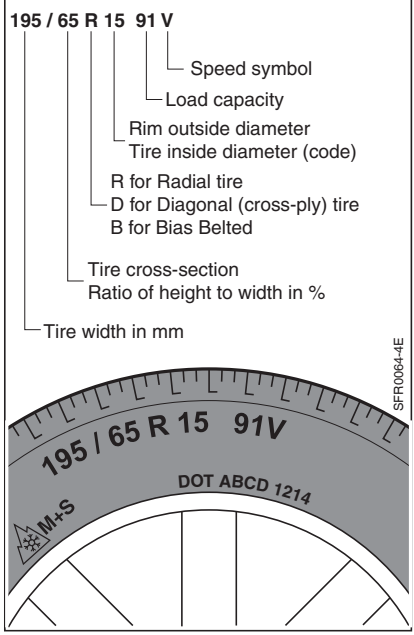
1 Rubber compound with distinct hysteresis (high tire-grip values),
2 Latest-generation rubber compound (combines low rolling resistance, good grip and high abrasion resistance),
3 Rubber compound with weak hysteresis (low rolling resistance, poor grip).



Rolling resistance and air pressure
Reduced tire pressure means increased flexing work and thus higher rolling resistance and impaired steering precision and braking stability.

Safety checks carried out in 2009 by the tire industry (Goodyear, Dunlop, Fulda) on 52,400 vehicles in 15 EU countries established that 81 % of all car drivers drive with excessively low tire pressure. Of these, 26.5 % were driving with clearly (down to 0.3 bar) and 7.5 % with significantly reduced pressure (0.75 bar and higher). The upshot of excessively low tire pressure is that fuel is wasted.

Figure 18: Tire designation (example)



Tire designation

Definition of terminology
According to EU Directives ECE 30 (for passenger cars, [5]), ECE 54 (for trucks, [6]) and ECE 75 (for motorcycles, [7]) tires must be provided with internationally agreed, standardized tire designations. This applies in particular to the tire sidewall (Figure 18). Tires tested in accordance with ECE (mandatory since October 1998) carry a burned-in circle with a capital “E” or a lower-case “e” and the code of the approving authority, e.g. E4, and followed by a release number (homologation number, see Figure 20).

The lettering, codes and symbols on the sidewall indicate, as well as the tire manufacturer and the type designation, the origin, the production date, the dimension, the load capacity, the maximum permissible speed, the tire design, and the ratio of tire width to tire height (tire cross-section).

Reading the information is made more difficult by the fact that units of measurement of the metric system standard in Central Europe (mm, bar) and of the British imperial system (1 inch = 25.4 mm) are jointly used.

The service description must feature near the size specification. This consists of the load index (LI, Table 1) and the speed symbol (SSY; also used is SI, speed index, Table 2) and provides information about the maximum load capacity of the tire in question for the top speed corresponding to the speed symbol. This regulation is binding in all EU member states and in Switzerland.

On tires with directional treads (often on winter and summer tires with V-shaped treads) an arrow on the sidewall indicates the prescribed rolling direction. If the tires are already mounted on the rim, they can only be mounted on one side of the vehicle and may not for example be exchanged cross-wise (diagonally) together with the rim.

Example: A 205/55 R 16 91 H size tire has a nominal width of 205 mm, the height of the sidewall is 55 % of the nominal width, which in this case in roughly 112 mm. The diameter of the matching rim to be mounted is 16 inches, i.e.

406 mm. The load capacity corresponds with a load index LI = 91 to the table value 615 kg – the vehicle axle load therefore according to the certificate of registration must not exceed 2 x 615 kg, i.e. 1,230 kg. With the speed index SI = H the vehicle may be driven with this tire at a speed not exceeding 210 km/h, even if the vehicle is designed to reach an higher final speed. Load-capacity reductions must be taken into account from speed index V and upwards.

Production date

There is usually a pressed-in four-digit number in an oval field on at least one of the two sidewalls next to the acronym DOT (US Department of Transportation) and a sequence of letters (code for manufacturing plant). This number denotes the production date. The first two digits indicate the calendar week, the last two digits the final digits of the production year (Figure 18).

Table 2: Speed index SI (speed symbol)

A1	up to 5 km/h	L	up to 120 km/h
A2	up to 10 km/h	M	up to 130 km/h
A3	up to 15 km/h	N	up to 140 km/h
A4	up to 20 km/h	P	up to 150 km/h
A5	up to 25 km/h	Q	up to 160 km/h
A6	up to 30 km/h	R	up to 170 km/h
A7	up to 35 km/h	S	up to 180 km/h
A8	up to 40 km/h	T	up to 190 km/h
B	up to 50 km/h	U	up to 200 km/h
C	up to 60 km/h	H	up to 210 km/h
D	up to 65 km/h	V	up to 240 km/h
F	up to 80 km/h	W	up to 270 km/h
G	up to 90 km/h	ZR	over 240 km/h
J	up to 100 km/h	Y	up to 300 km/h
K	up to 110 km/h	(Y)	over 300 km/h

Table 1: Load index LI (values up to 3,350 kg, table open in upward direction)

LI	kg	LI	kg	LI	kg	LI	kg	LI	kg	LI	kg
1	46.2	26	95	51	195	76	400	101	825	126	1,700
2	47.5	27	97.5	52	200	77	412	102	850	127	1,750
3	48.7	28	100	53	206	78	425	103	875	128	1,800
4	50	29	103	54	212	79	437	104	900	129	1,850
5	51.5	30	106	55	218	80	450	105	925	130	1,900
6	53	31	109	56	224	81	462	106	950	131	1,950
7	54.5	32	112	57	230	82	475	107	1,000	132	2,000
8	56	33	115	58	236	83	487	108	1,030	133	2,060
9	58	34	118	59	243	84	500	109	1,060	134	2,120
10	60	35	121	60	250	85	515	110	1,090	135	2,180
11	61.5	36	125	61	257	86	530	111	1,120	136	2,240
12	63	37	128	62	265	87	545	112	1,150	137	2,300
13	65	38	132	63	272	88	560	113	1,180	138	2,360
14	67	39	136	64	280	89	580	114	1,215	139	2,430
15	69	40	140	65	290	90	600	115	1,250	140	2,500
16	71	41	145	66	300	91	615	116	1,285	141	2,575
17	73	42	150	67	307	92	630	117	1,320	142	2,650
18	75	43	155	68	315	93	650	118	1,360	143	2,725
19	77.5	44	160	69	325	94	670	119	1,400	144	2,800
20	80	45	165	70	335	95	690	120	1,450	145	2,900
21	82.5	46	170	71	345	96	710	121	1,500	146	3,000
22	85	47	175	72	355	97	730	122	1,550	147	3,075
23	87.5	48	180	73	365	98	750	123	1,600	148	3,150
24	90	49	185	74	375	99	775	124	1,650	149	3,250
25	92.5	50	190	75	387	100	800	125	1,700	150	3,350

Example: 1214 means the 12th week of the year 2014. Prior to the year 2000 the designation of the production date had only three digits.

Special case:
Winter-tire designation
M+S tires
Winter tires must carry the M+S symbol (Mud and Snow) (see Figure 18). EU Regulation No. 661/2009 [8] denotes as an M+S tire a tire whose tread pattern, tread compound or construction is designed first and foremost "to achieve compared with a summer tire better values for winter handling performance and traction on snow" – an extremely woolly definition.

To compare: The previously applicable EU Regulation from 1992 [9] stated that "M+S tires" are such tires "on which the pattern of the tread and the structure are designed in such a way that they guarantee above all in mud and fresh or melting snow better handling performance than normal tires. The pattern of the tread of M+S tires is generally characterized by larger tread grooves and lugs which are separated from each other by larger spaces than is the case on normal tires".

"M+S" is to date not a protected or precisely defined designation and for this reason may also be used on tires not suitable for winter driving conditions (i.e. also on summer tires). The M+S symbol no longer has any significance with regard to suitability for winter driving conditions.

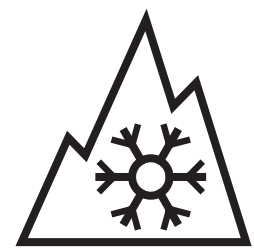
Snow-flake symbol
A revision of the M+S designation was called for in response to the widespread traffic chaos in the USA in 1995. A winter tire should satisfy certain criteria with regard to its suitability for winter driving conditions and verify this by way of appropriate tests. This gave rise to the designation 3PMSF (Three Peak Mountain Snow Flake), which today is firmly established in North American legislation.

This designation has also be used voluntarily for a few years in Europe. This is intended to verify to the consumer suitability for winter driving conditions substantiated by tests. The test criteria are defined by the European Union in UN-ECE R 117 [11].

Amendment of StVZO §36 section 4 [12]
With the 52nd Regulation to amend road-traffic legislation of 18 May 2017 the definition of tires for winter weather conditions was adapted and consequently a new winter-tire regulation came into force. By force of this regulation the snow-flake symbol for winter tires is already mandatory for the winter season 2017/2018 in Germany and consequently replaces the M+S marking. In a transition phase lasting until 30 September 2024 tires which only have an M+S marking which continue to be classed as winter tires provided there date of manufacture is before 31 December 2017. M+S tires which are manufactured from 2018 onwards and only have an M+S marking will no longer be classed as winter tires.

For truck and bus tires on the steering axle the introduction of this regulation will be deferred until no later than 1 July 2020. This time will be used to check as part of a study the necessity of the new winter-tire regulation for the steering axle of these vehicle categories. The deferment of the winter-tire regulation does not apply to the permanently driven axles of these vehicles.

Figure 19: Snow-flake symbol (3PMSF) in acc. with UN-ECE R117



Sound identification
Tires featuring this marking comply with Directive ECE 2001/43 [10], which lays down the maximum values for rolling noise. This marking has been mandatory since 1 October 2009, is featured next to the UN/ECE approval mark, and can be recognized by the code letter “s” after the homologation number.

With the introduction of the tire label on 1 November 2012 this designation for new tire types was extended by a capital “S” and by the letters “W” for wet grip and “R” for rolling resistance (Figure 20). Additional code numbers (1 and 2) after the letters S and R denote the limit values which must be adhered to for the prescribed time phases (periods) (regulations, see UN/ECE Regulation No. 117 [11]).

Figure 20: E designation
Example of a homologation number acc. to ECE R 117 for complying with rolling noise.



EU tire label

Definition of terminology
Since November 2012 new tires sold in the EU which were produced from July 2012 onwards must be provided with a standardized tire label (7.5 cm × 11 cm) (Figure 21). This label is meant to provide the buyer with quick and unmistakable information on the three tire properties of rolling resistance, wet grip (restricted to wet braking distance) and pass-by noise (not passenger compartment), thereby helping them to make a more informed decision to purchase. The winter properties of winter tires are currently not recorded by the EU tire label.

The tire label records tires in the categories C1 (passenger cars), C2 (light utility vehicles) and C3 (heavy commercial vehicles) as set out in Table 3. Excluded from the regulation are retreaded tires,

Figure 21: EU label

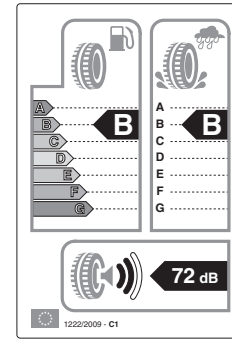


Table 3: Energy-efficiency classes of the different tire categories

Class C1 tires (passenger cars)		Class C2 tires (light utility vehicles)		Class C3 tires (heavy commercial vehicles)	
c_{RR} in kg/t	Energy-efficiency class	c_{RR} in kg/t	Energy-efficiency class	c_{RR} in kg/t	Energy-efficiency class
$c_{RR} \le 6.5$	A	$c_{RR} \le 5.5$	A	$c_{RR} \le 4.0$	A
$6.6 \le c_{RR} \le 7.7$	B	$5.6 \le c_{RR} \le 6.7$	B	$4.1 \le c_{RR} \le 5.0$	B
$7.8 \le c_{RR} \le 9.0$	C	$6.8 \le c_{RR} \le 8.0$	C	$5.1 \le c_{RR} \le 6.0$	C
not assigned	D	not assigned	D	$6.1 \le c_{RR} \le 7.0$	D
$9.1 \le c_{RR} \le 10.5$	E	$8.1 \le c_{RR} \le 9.2$	E	$7.1 \le c_{RR} \le 8.0$	E
$10.6 \le c_{RR} \le 12.0$	F	$9.3 \le c_{RR} \le 10.5$	F	$c_{RR} \ge 8.1$	F
$c_{RR} \ge 12.1$	G	$c_{RR} \ge 10.6$	G		
Example: c_{RR} = 10.5 kg/t (label class E) corresponds to c_{RR} = 0.0105 or F_{RR} = 105 N.					

professional off-road tires, racing tires, spikes, compact spare-wheel tires, vintage-car and modern classic-car tires (for vehicles first registered before 1 October 1990), tires for top speeds below 80 km/h, tires with inside diameters of less than 254 mm or more than 635 mm, and motorcycle tires.

Naturally, the tire label cannot show all the tire criteria (there are up to fifty), but the criteria selected do represent a certain combination of many other linked properties.

Fuel consumption
Letters from A (highest efficiency) through G (lowest efficiency) and the traffic-light colors green, yellow and red in the EU label denote the efficiency of the tire with regard to rolling resistance and thus fuel consumption. The range for rolling resistance from class A through G represents a difference in fuel consumption of up to 7.5 % [13]. The potential for savings is even higher, depending on the driving situation. The difference between the individual stages of the tire classes with regard to rolling resistance is clearly defined: for instance 0.11 *l* / 100 km for a vehicle with an average consumption of approx. 6.6 *l* / 100 km. The difference between an A-rated tire and a G-rated tire adds up to approx. 0.5 *l* / 100 km (in each case provided a consistent driving style and identical tire inflation pressure).

Wet grip
Letters A (shortest braking distance) through G (longest braking distance) provide information on the tire's wet grip when braking. The difference in wet grip between a particularly good tire and a poor tire when applying full brakes from 80 km/h to zero produces an 18-meter shorter braking distance.

Tire noise
Pass-by noise is depicted in the EU label by a symbol together with an indication of the dB value. The symbol is based on the binding noise-emission limit values. The more sound waves are shaded black in the EU label, the louder the tire.

Winter tires

Technical characteristics
Passenger-car winter tires exhibit no or only slight differences in structural design from summer tires. They are characterized by good force transmission (traction) on snow and mud, satisfactory grip on ice, good adhesion on wet and dry road surfaces, safe handling, comfortable rolling, and low noise.

The feature that is unique to them is primarily the softer and elastic-when-cold rubber compound with high natural-rubber content. Unlike the summer rubber compound, it does not become brittle and harden ("vitrify") at minus temperatures (which results in reduced adhesion because the interlocking effect can no longer occur).

Added to this are increased tread area (thus a greater positive proportion) and, as a conspicuous external identifying feature, a multidimensional superfine tread (design shapes: zigzag, spheres, honeycombs, and mixed shapes, see Figure 4) in the rubber blocks themselves. These sipes (up to 2,000, depending on tire size) offer additional gripping edges in the snow and increase both traction and braking performance noticeably.

As the tread depth decreases, but also with increasing age (hardening caused by among others the influence of UV and ozone) the winter properties diminish: Traction and cornering-stability potential deteriorate, the braking distance is extended, and the risk of aquaplaning increases. Winter tires identified for example in Austria under a remaining tread depth of 4 mm are always classed as summer tires.

Standard winter tires customary in Europe are offered in speed ratings up to speed index W (maximum speed 270 km/h).

Development of a tire

The tire is always a compromise product: If a particular property (e.g. rolling resistance) is predominantly developed, this is inevitably at the expense of other properties (e.g. with the upshot of reduced wet grip) and thus of balance. This is only desirable or permissible in special cases (motor sport, special tires, industrial requirements). There is a conflict of aims when on principle contrasting properties are to be simultaneously optimized. Modern rubber compounds with a high silica filler content and optimized ground contact patches raise this balancing act to a higher level, but this does not eliminate the conflict.

The conflicting development aims which can negatively influence each other include grip potential and rolling resistance, dry braking and wet braking, aquaplaning and dry handling, and grip and abrasion.

Tire tests

Tire manufacturers conduct around fifty objective laboratory tests and subjective road tests on in-house test and racing circuits. The most important tests are:

Handling on a dry road
Criteria are directional stability, steering precision, straight-running stability, noise, and tire comfort.

Properties in the wet
Criteria are handling on a wet course with bends, braking, aquaplaning, longitudinal and lateral, and circular-course driving.

Machine tests
Criteria are top speed, continuous running, and abrasion.

Winter properties
Criteria are snow handling, driving on passes, traction measurement, accelerating performance, and braking.

References for tires

[1] Source: Goodyear Dunlop, 2012.
[2] Council Directive 89/459/EEC of 18 July 1989 on the approximation of the laws of the Member States relating to the tread depth of tyres of certain categories of motor vehicles and their trailers.
[3] Source: Continental AG, 2011. The specified brake differences were determined with a Mercedes C-Class car on 205/55 R 16 V size tyres in over 1,000 brake tests.
[4] Source: Michelin tire plants, 2010.
[5] ECE 30: Regulation No. 30 of the United Nations Economic Commission for Europe (UN/ECE) – Uniform provisions concerning the approval of pneumatic tyres for motor vehicles and their trailers.
[6] ECE 54: Regulation No. 54 – Uniform provisions concerning the approval of pneumatic tyres for commercial vehicles and their trailers.
[7] ECE 75: Regulation No. 75 – Uniform provisions concerning the approval of pneumatic tyres for motor cycles and mopeds.
[8] Regulation (EC) No. 661/2009 of the European Parliament and of the Council of 13 July 2009 concerning type-approval requirements for the general safety of motor vehicles, their trailers and systems, components and separate technical units intended therefore.
[9] Council Directive 92/23/EEC of 31 March 1992 relating to tyres for motor vehicles and their trailers and to their fitting installation.
[10] Directive 2001/43/EC of the European Parliament and of the Council of 27 June 2001 amending Council Directive 92/23/EEC relating to tyres for motor vehicles and their trailers and to their fitting.
[11] Regulation No. 117 of the United Nations Economic Commission for Europe (UN/ECE) – Uniform provisions concerning the approval of tyres with regard to rolling sound emissions and/or to adhesion on wet surfaces and/or to rolling resistance.
[12] §36 StVZO: Bereifung, Laufflächen (Tires, treads).
[13] Source: Tire manufacturers.

Tire-pressure monitoring systems

Application

Tire-Pressure Monitoring Systems (TPMS) are used to monitor the tire pressure on vehicles to prevent tire defects due to insufficient tire pressure, thus reducing the number of accidents resulting from defective tires.

If a vehicle is operated with insufficient tire pressure, this leads to increased flexing energy on the tire sidewalls and thus to increased wear of the tire. When operated at full load or at high speed, the greater flexing energy results in increased thermal load, which can even cause tire bursts. Following a spate of severe fatal accidents in the USA due to tire bursts caused by insufficient inflation pressure, legislation was passed (NHTSA Tread Act) to regulate the nationwide introduction of tire-pressure monitoring systems in the USA in order to warn drivers about low tire inflation pressure at an early stage in the future. Since September 2007, all new cars have been required to be fitted with tire-pressure monitoring systems which detect both tire damage and slow pressure losses through the tire rubber caused by gas diffusion.

The tire inflation pressure, however, is not only an important variable for traffic safety. Ride comfort, tire service life, and fuel consumption are also significantly influenced by the inflation pressure. Inflation pressure reduced by 0.6 bar can increase fuel consumption by up to 4 % in urban traffic and shorten the service life of the tire by up to 50 %. In the European Union (EU), the decision has been taken to prescribe the fitting of tire-pressure monitoring systems in all new cars as from October 2012 in order to help reduce CO_2 emissions.

Already today the rising proportion of tires with run-flat properties necessitates the deployment of tire-pressure monitoring systems, as the car driver is no longer able to detect a tire with a considerable pressure deficiency (“flat”) on the basis of the drivability. In order to prevent the driver from inadvertently exceeding the speed and range limits that apply in this case, run-flat tires may only be used in conjunction with tire-pressure monitoring systems.

As a general principle, two different types of tire-pressure monitoring system are used: directly measuring and indirectly measuring systems.

Directly measuring systems

In directly measuring systems, a sensor module with a pressure sensor is installed in each tire of the vehicle. This transfers data such as the tire pressure and tire temperature from inside the tire across a coded high-frequency transmission link to a control unit. The control unit evaluates these data in order to detect not only pressure losses in individual tires (“puncture detection”) but also slow pressure losses in all the tires (“diffusion detection”). If the tire pressure falls below a specified threshold or if the pressure gradient exceeds a certain value, the driver is warned by a visual or acoustic signal.

The sensor modules are usually integrated into the tire valve. As a rule, they are supplied by a battery. In comparison with other applications, this results in additional requirements with regard to power consumption, media resistance, and sensitivity to acceleration. Micromechanical absolute-pressure sensors are used as sensor elements.

The data measured with the pressure and temperature sensor in the tire are processed in the sensor module, modulated on a HF carrier signal (433 MHz in Europe, 315 MHz in the USA), and emitted via an antenna. This signal is either detected via individual antennas on the wheel arches or in a central receiver (e.g. in the control unit of existing Remote Keyless Entry systems).

Directly measuring systems do not need a reset function if they have a fixed, constant pressure-loss warning threshold. For such a system to work, the vehicle must only have one prescribed inflation pressure, regardless of vehicle load and tire size. As soon as different inflation pressures have to be set in the vehicle, a directly measuring system also needs a reset function to be able to adapt the warning threshold accordingly.

The advantages of directly measuring systems are that they provide precise,

real measurement of the tire pressure and temperature, and their functioning is not dependent on specific tire types, vehicle conditions and road conditions. The disadvantages of direct systems compared with indirect systems are the much higher system costs, the additional logistical costs in the field of maintaining the availability of all the design variants, the follow-up costs for each new rim, and their battery-dependent, limited service life.

Indirectly measuring systems

In indirectly measuring systems, pressure loss in the tires is not determined directly, but rather by means of a derived variable. To achieve this, these systems perform a mathematical-statistical evaluation of the speed differences of all wheels for "puncture detection", and if necessary also an evaluation of a wheel natural-frequency shift for "diffusion detection". In vehicles with antilock braking or driving-dynamics control systems, the wheel speed required for this is determined by sensors that are already present and transferred to the control unit. Speed differences occur when pressure loss reduces the rolling circumference of the corresponding tire, thus increasing its speed relative to the other three wheels. The subtraction, which can be implemented using a low-cost extension of the software algorithms in the antilock braking or driving-dynamics control system, enables detection of high pressure losses on up to three tires. The wheel natural-frequency spectrum of the individual wheels is evaluated to enable a simultaneous pressure loss at all four wheels to be detected. Typically, the maximum wheel natural frequency shifts at a pressure loss of 20% from 40 Hz to approximately 38 Hz.

Indirectly measuring systems must necessarily be calibrated to the nominal pressure. A calibration is initiated by operating the reset button. When the reset function is activated, the system stores the current learning values on the next few kilometers as new reference values, based on the current rolling circumferences and wheel natural-frequency characteristics. The warning capability takes effect after approx. ten minutes of driving time. The driver is required to activate the reset function to recalibrate the system when one or more tires are changed, the tire positions are changed (e.g. switching the front and rear wheels), the tire pressure is altered (e.g. when the vehicle is fully laden), or work has been carried out on the wheel suspension (e.g. adjustment work, shock-absorber replacement).

Advantages of indirectly measuring systems are their lower system costs and their robustness over the service life of vehicles, since no additional components are required. Because the system is linked to the vehicle and not to the wheels, no further costs are incurred in the field for logistics and spare parts. A disadvantage is the system's dependence on the specific tire, resulting in a wider variation of the detection times and in higher costs for adapting the system's function to the tire dimensions permitted for a particular vehicle. The system's dependence on mileage and road surface also influences the detection times.

Fulfillment of statutory requirements

As things stand, both systems satisfy the statutory requirements in North America and Europe with regard to "puncture detection" and "diffusion detection". Statutory provisions for tire-pressure monitoring systems are also being drafted in China and Korea.

Rotary seal for tire-pressure control

Function

The Turcon PTFE rotary seal (Turcon Roto L, Figure 1) was developed for tire-pressure control purposes in order to seal around the axle of the central tire-pressure system only when required and always when the tire pressure is increased or decreased. In conventional seal concepts the seal remains in permanent contact with the axle that bears the seal. This causes friction, which in turn leads to increased fuel consumption. By avoiding friction in pressureless operation this rotary seal necessarily reduces fuel consumption.

Application

The Turcon Roto L was originally developed primarily for off-road vehicles to adapt the tire inflation pressure to the ground. On asphalt surfaces a high tire inflation pressure is required, whereas on dirt tracks a low tire inflation pressure is advantageous.

The use of this tire-pressure control can also produce benefits for trucks. It facilitates the extension of tire-pressure control systems and produces benefits for trucks with long loading areas or different semi-trailers in which the tire pressure could be optimized for the respective road and load conditions. This would increase traction and safety and help to deliver significant fuel savings to transport companies.

Performance and passenger cars are a further target market. When sports cars are driving at low speeds the tire pressure should be kept low. At high speeds the handling performance is improved if the tire pressure can be set to high or, depending on the road surface, set differently at each wheel. Furthermore, it can be used as a safety system in the event of a blow-out in order to inflate the system with air until the driver reaches the nearest repair shop.

Design

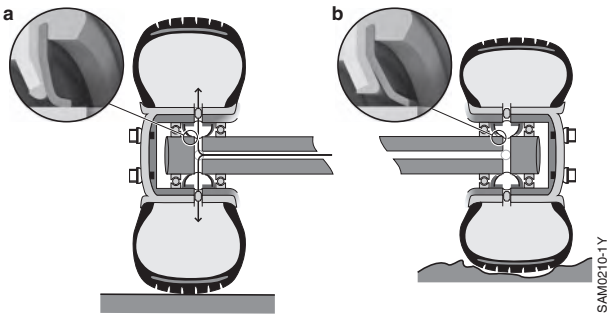
The Turcon Roto L combines a polytetrafluoroethylene (PTFE) sealing lip with an elastomer seal body and a stable, shape-forming metal ring (Figure 2). The seal's design ensures that during pressure build-up the sealing lip is pressed against the mating sealing surface. As the pressure decreases, the relaxing elastomer area, which acts like a spring, returns the sealing lip to its neutral initial position.

Seal wear

A further important effect on the performance of the Turcon Roto L with regard to the service life of the overall system is the significantly reduced shaft run-in. This is due on the one hand to the use of a special seal material which is filled with a complex mixture of non-abrasive mineral fibers and on the other hand to the low friction when the seal is not activated.

A conventional standard seal used today in a tire-pressure control system already exhibits a shaft run-in of around 10 µm after 168 hours. In comparison, the Turcon Roto L has a shaft run-in of around 4 µm after 780 hours (Figure 3). In other words, the service life of a sealing system with this rotary seal is at least four times longer.

Figure 1: Turcon Roto L sealing system
a) Setting a high tire inflation pressure,
b) Setting a low tire inflation pressure.



Seal seat
Rotary seals for tire-pressure control need to be well seated in the groove to ensure tightness, even when the system is pressurized. In this way, a relative movement of the body is also prevented when the shaft is subject to axial movement. The groove should have a diameter tolerance of H8 (according to DIN 3760 [1], ISO 6194-1 [2], and ISO 16589-1 [3]). It is important that the seal has not been subjected to load by other components beforehand (for example, by bearings).

The shaft should have a diameter tolerance of h7. The surface should be ground to a perfectly flat finish. A surface with Ra = 0.2 to 0.4 µm and a maximum Rz value of 1.5 µm is recommended for this rotary seal.

Benefits
One benefit of this concept is that it extends the seal life. If the seal in a vehicle is only used for around 10 % of the axle system's service life, it remains operational and fit for use over the system's entire service life.

Even more important, however, is the elimination of friction when there is no pressurization. This means considerable fuel savings for the vehicle operator. Because the seal generates less than half the friction of conventional seal solutions, the tire pressure can also be reduced or increased during driving. In the case of large commercial-vehicle tires it can take 20 to 30 minutes to adjust the tire pressure. If this can be done en route while the vehicle is driving, the vehicle can be active for up to 30 minutes instead of standing idle. This means a reduction of overall costs for the operator.

References for rotary seal for tire-pressure control
[1] DIN 3760: Rotary shaft lip type seals; 1996.
[2] ISO 6194-1: Rotary shaft lip-type seals incorporating elastomeric sealing elements – Part 1: Nominal dimensions and tolerances; 2007.
[3] ISO 16589-1: Rotary shaft lip-type seals incorporating thermoplastic sealing elements – Part 1: Nominal dimensions and tolerances; 2011.

Figure 2: Turcon Roto L design
1 Metal ring,
2 Sealing lip.

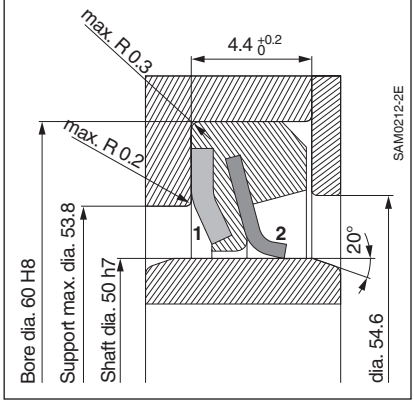
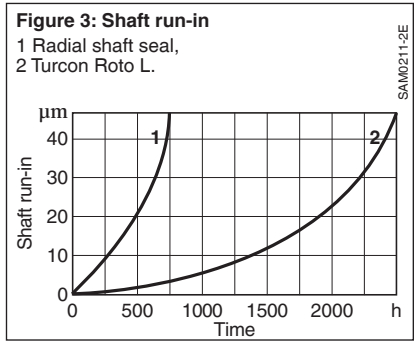


Figure 3: Shaft run-in
1 Radial shaft seal,
2 Turcon Roto L.



Steering

Purpose of automotive steering systems

The direction of vehicles of any kind is changed with the steering. In single-track vehicles, for example motorcycles, steering is performed by turning the pivoted front fork in the steering head (steering-head steering). In two-track vehicles kingpin steering as described in the following is used.

In road vehicles the wheels on the front axle are steered; in some vehicle models the wheels on the rear axle also steer. Rear-axle steering reduces the turning circle at low speeds and increases the driving dynamics at high speeds.

Kingpin steering

In a kingpin-steering system the two wheels are each secured to a pivoted steering knuckle (Figure 1). The rotary movement is introduced via the steering arms rigidly connected to the steering knuckle by the two tie rods actuated by the steering box. The casing of the steering box is permanently connected to the vehicle. The rotary movement is introduced via the steering column, at the upper end of which the steering wheel is located.

Aside from the vehicle weight and the friction coefficient of the wheels on the road surface, the axle kinematics affect the steering behavior and the forces to be applied by the steering. As a result of the toe-in (see Fundamentals of automotive engineering) the two front wheels are not parallel but instead are slightly inclined towards each other in the direction of travel in order to improve straight-running stability. As a result of the camber the wheels are not exactly vertical on the road but instead are slightly inclined inwards. This results in a side force on the wheels acting towards the inside of the vehicle. This force pretensions the bearings in the straight-ahead position and provides a better steering response and greater directional stability. The kingpin inclination influences the steering forces. In the transverse direction the swivel axis of the steering knuckle is inclined inwards, equating to a positive kingpin angle. A small kingpin angle reduces the steering

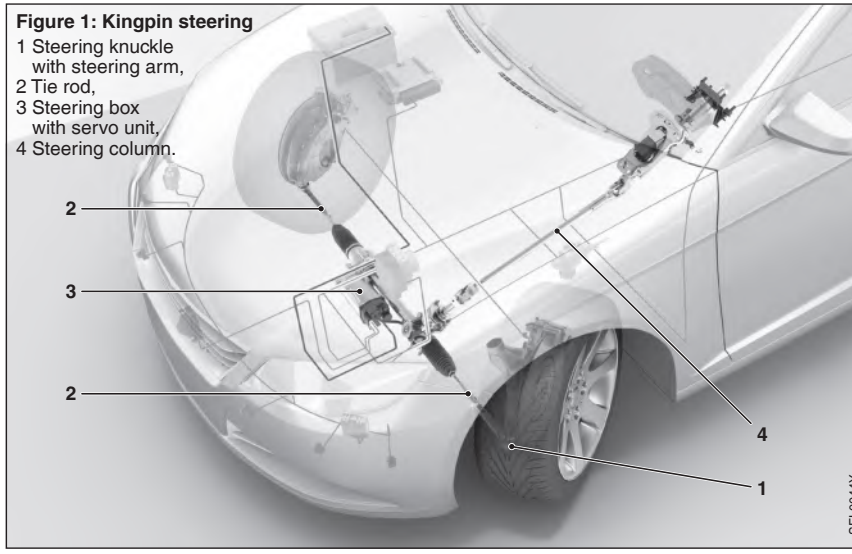


Figure 1: Kingpin steering
1 Steering knuckle with steering arm,
2 Tie rod,
3 Steering box with servo unit,
4 Steering column.

forces, whereas a large kingpin inclination assists the automatic return to the straight-ahead position.

The automatic return to the straight-ahead position is achieved above all by the caster offset. Inclining the swivel axes in the direction of travel towards the rear moves the piercing point of the swivel axis through the road surface in the direction of travel ahead of the wheel contact point and consequently the wheels are pulled in the direction of travel.

Classification of automotive steering systems

Muscular-energy steering system
The required steering forces are generated exclusively by the muscular energy of the driver. These steering systems are currently used in the smallest vehicle categories.

Power-assisted steering system
The steering forces are generated by the muscular energy of the driver and by an additional auxiliary force hydraulically and increasingly electrically. This steering system is currently the type typically used in passenger cars and commercial vehicles.

Power-steering system
The steering forces are generated exclusively by non-muscular (external) energy (e.g., in machinery).

Friction steering system
The steering forces are generated by forces which act on the tire contact patch. The trailing axles in trucks are an example of this type. The transmissions for the steering and auxiliary forces are effected by mechanical, hydraulic or electrical means or by combinations of the three.

Steering-system requirements

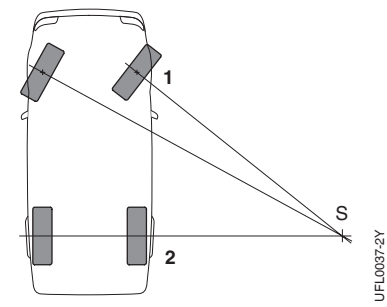
General requirements
To ensure that the road wheels roll cleanly and are thereby not subjected to excessive tire wear, the entire steering kinematics must satisfy the Ackermann condition. This means that the extensions of the wheel axes of the steered wheels intersect at the same point on the extension of the wheel axis of the non-steered wheels (Figure 2).

Steering kinematics and axle design must be such that, although the driver receives feedback on the adhesion between wheels and road surface, the steering wheel is not subjected if possible to any forces from the spring motion of the wheels or from motive forces (front-wheel drive). Furthermore, damping properties must be utilized to ensure that no vibration, caused by road irregularities for exampled, occurs in the steering train.

Jolts from irregularities in the road surface should be damped as much as possible during transmission to the steering wheel. But, in the process, the required haptic feedback from the road to the driver must not be lost.

The steering-angle requirement for turning the steering wheel from lock to lock should, for comfort reasons, be as small as possible when parking and driving at low speed. However, the direct

Figure 2: Ackermann condition
1 Front axle,
2 Rear axle.
S Point of intersection of extension of wheel axes.



steering ratio should not make the vehicle susceptible to instability at medium and high speeds.

Statutory requirements

Displacement and time synchronization when the steering system is operated is required by law. A rigid, zero-play configuration of all the components in the steering train helps to achieve this and causes the slightest steering-wheel movements by the driver to be translated into changes of direction at the steered wheel. This property is particularly important in the range of the straight-ahead position. This makes possible safe, precise, and fatigue-free vehicle guidance.

A further statutory requirement is the tendency of the steering to return automatically to the straight-ahead position. This is achieved by appropriate axle kinematics and a low-friction design of the components in the steering train.

The statutory requirements imposed on steering systems in motor vehicles are described in the international regulation ECE-R79 [1]. These requirements include, as well as the basic functional requirements, the maximum permissible control forces for an intact steering system and for a faulty steering system. These requirements govern, above all, the behavior of the vehicle and the steering system when driving into and out of a circle. Owing to different requirements, vehicles are divided in ECE-R79 into different categories, to which different stipulations then apply (Table 1).

For vehicles of all categories: After the steering wheel is released when the vehicle is driven on a circular course at half lock and at a speed of 10 km/h or faster, the driven radius of the vehicle must become larger or at least remain the same.

For M1 category vehicles (passenger cars with up to eight seats in addition

Table 1: Categorization of vehicles in ECE-R79

Vehicle category	
M1	Vehicles for the carriage of passengers comprising no more than 8 seats in addition to the driver's seat
M2	Vehicles for the carriage of passengers comprising more than 8 seats in addition to the driver's seat and having a maximum mass not exceeding 5 tons
M3	Vehicles for the carriage of passengers comprising more than 8 seats in addition to the driver's seat and having a maximum mass exceeding 5 tons
N1	Vehicles for the carriage of goods and having a maximum mass not exceeding 3.5 tons
N2	Vehicles for the carriage of goods and having a maximum mass exceeding 3.5 tons but not exceeding 12 tons
N3	Vehicles for the carriage of goods and having a maximum mass exceeding 12 tons

Table 2: Regulations for steering operating force when driving into a circle at a speed of 10 km/h

Vehicle category	Intact system			Faulty system		
	Maximum operating force in daN[1]	Time in s	Turning circle radius in m	Maximum operating force in daN[1]	Time in s	Turning circle radius in m
M1	15	4	12	30	4	20
M2	15	4	12	30	4	20
M3	20	4	12 [2]	45	6	20
N1	20	4	12	30	4	20
N2	25	4	12	40	4	20
N3	20	4	12 [2]	45 [3]	6	20

[1] 1 daN = 10 N.
[2] Or steering lock in case this value is not reached.
[3] 50 daN for non-articulated vehicles, with two or more steered axles, excluding friction-steered axles.

to the driver's seat): When the vehicle is driven tangentially out of a circle with a radius of 50 m at a speed of 50 km/h, no unusual vibrations may occur in the steering system. In vehicles of the other categories, this behavior must be substantiated at a speed of 40 km/h or, if this value is not reached, at top speed.

This behavior is also prescribed in the event of a fault in vehicles with power-assisted steering systems. For M1 category vehicles, it must be possible in the event of a steering-servo failure to drive at a speed of 10 km/h within four seconds into a circle with a radius of 20 m. The control force at the steering wheel must not exceed 30 daN in the process (Table 2). The same applies to the behavior with an intact system.

Types of steering box

The specified steering-system requirements have given rise above all to two fundamental types of steering box. Both types can be utilized in pure muscular-energy steering systems or, in combination with appropriate servo systems, as power-assisted steering systems.

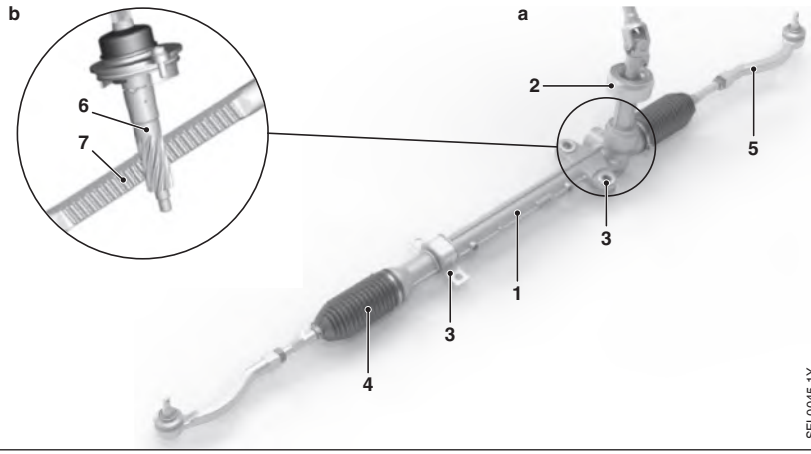
Rack-and-pinion steering

Rack-and-pinion steering is by far the most commonly used type and consists essentially of steering pinion and rack (Figure 3). The steering pinion is driven by the steering wheel and the steering column (Figure 1). The steering ratio is defined by the ratio of pinion revolutions (steering-wheel revolutions) to rack travel.

As an alternative to a constant reduction ratio on the rack, suitable toothing of the rack allows the ratio to be varied as a function of travel. In this way, the straight-running stability of the vehicle can be improved by a suitably indirect

Figure 3: Rack-and-pinion steering

a) Design,
b) Detail.
1 Steering-gear box with rack,
2 Steering spindle with pinion,
3 Steering mounting,
4 Gaiter,
5 Tie rods with outer links,
6 Steering pinion,
7 Rack with toothing.

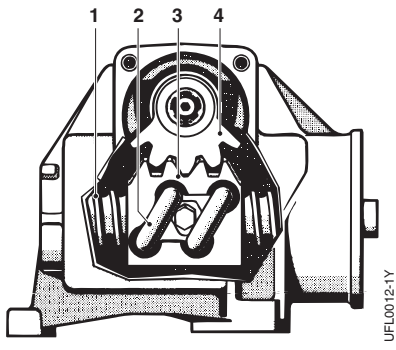


ratio around the center of the steering. At the same time, it is possible with a direct ratio arrangement in the range of medium and large steering angles (e.g., when parking) to reduce the necessary steering-angle requirement when turning from lock to lock.

Recirculating-ball steering
The forces generated between steering worm and steering nut are transmitted via a low-friction row of recirculating balls (Figure 4). The steering nut acts on the steering output shaft via gear teeth. A variable ratio is also possible with this type of steering box.

The increasing performance of rack-and-pinion steering has meant that recirculating-ball steering is practically no longer used in passenger cars.

Figure 4: Recirculating-ball steering
1 Steering worm,
2 Recirculating balls,
3 Steering nut,
4 Steering output shaft with toothing segment.



Power-assisted steering systems for passenger cars

The upshot of the increasing size and weight of vehicles and the heightened comfort and safety requirements involved is that in the last few years power-assisted steering has gained acceptance in all vehicle categories down to the compact-car segment. These steering systems are, but for a few exceptions, already installed as standard. The steering forces exerted by the drive are boosted by a hydraulic or an electric servo system. This servo system must be such that the driver receives good feedback on the adhesion conditions between the tires and the road surface at all times, and yet negative influences caused by road-surface jolts are effectively damped.

Hydraulic power-assisted steering
Combining the mechanical type of steering box with a hydraulic servo system produces rack-and-pinion power steering (Figure 6) and ball-and-nut power steering.

Control valve
The control valve provides the working cylinder with an oil pressure that corresponds to the rotary force of the steering wheel (Figure 5). For this purpose, a torsion-bar spring is used to convert the operating torque into proportional actuator travel inside the valve. The actuator travel changes opening cross-sections inside the valve and thereby control the oil and pressure ratios in the working cylinder.

Control valves are built according to the "open center" principle, i.e., when the control valve is not actuated, the oil delivered by the pump flows back to the oil reservoir at zero pressure.

Power-assisted steering systems with modulation capability
Increasing demands regarding user-friendliness and safety have resulted in the introduction of power-assisted steering systems with modulation capability. An example of this is the rack-and-pinion power-steering system operating in response to driving speed. An ECU evaluates the driving speed and controls,

via an electrohydraulic control valve, the hydraulic feedback in the system and consequently the operating force at the steering wheel.

This produces in the parking range and when driving slowly low operating forces, which then increase as the speed increases. In this way, precise and accurate steering is possible at high speeds.

Working cylinder

The double-action working cylinder integrated in the steering box converts the applied oil pressure into an assisting force which acts on the rack and intensifies the steering force exerted by the driver. As the working cylinder has to have extremely low friction, particularly high demands are made on the piston and rod seals.

Figure 5: Functioning principle of control valve of hydraulic power-assisted steering

a) Control valve in neutral position,
b) Control valve in working position.
1 Power-steering pump,
2 Control bushing,
3 Rotary slide,
4 Left working-cylinder chamber,
5 Right working-cylinder chamber,
6 Oil reservoir.

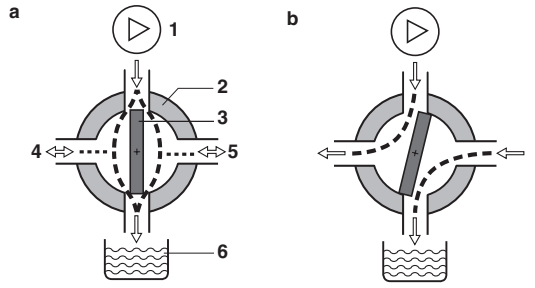
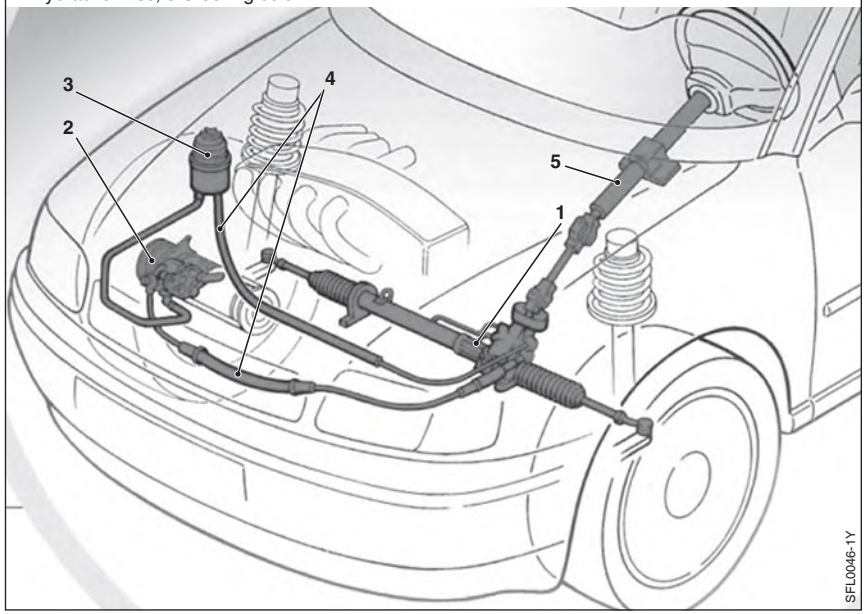


Figure 6: Rack-and-pinion power steering

1 Hydraulic steering box, 2 Power-steering pump, 3 Oil reservoir,
4 Hydraulic lines, 5 Steering column.



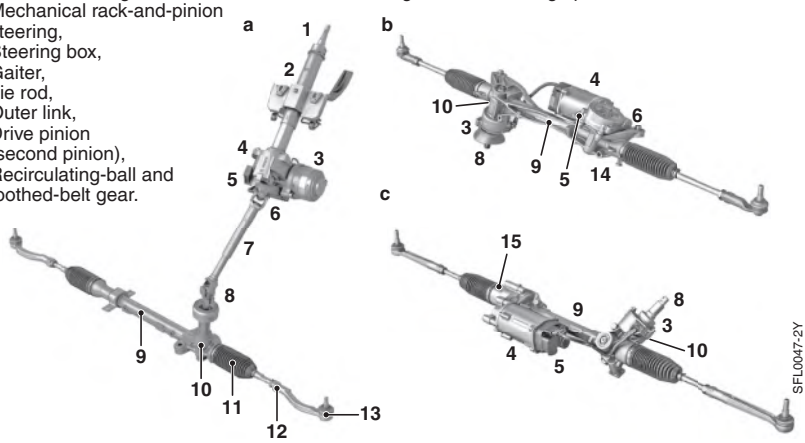
Energy source
The energy source consists of a vane-type supply pump (generally driven by the internal-combustion engine) (power-steering pump) with an integrated oil-flow regulator, an oil reservoir, and connecting hoses and pipes. The pump must be dimensioned so that it makes available the required oil pressure and the required oil quantity for parking maneuvers, even when the engine is just idling.

To provide protection against overload, steering systems contain a pressure-relief valve which is usually integrated in the pump. The pump and the components in the hydraulic working circuit must be designed so that no offending noise is generated and the operating temperature of the hydraulic fluid does not rise to an excessive level.

Alternatively, the pump for the energy supply of the steering can also be driven by an electric motor. The pump is usually designed as a gear or roller-cell pump. The pump can be mounted in a variety of locations due to the omission of the belt drive for the internal-combustion engine. Thanks to an electronic control module which evaluates signals like the vehicle or steering speed, the pump's rotational speed can be adapted to the steering's current energy demand and to the driving situation, and consequently greater energy efficiency can be achieved.

Electric power-assisted steering
Hydraulic power-assisted steering has been replaced in the passenger-car sector in the last few years almost exclusively by electromechanical power-assisted steering (EPS, Electric Power Steering). Introduced in micro and compact cars in around 1990 and then in the compact class in around 2000, ever more powerful electric ECUs and motors enable such systems to be used even in the upper passenger-car and SUV sectors. Electromechanical power-assisted steering is characterized by an electric drive unit consisting of an ECU and an electric motor powered by the vehicle electrical system which delivers the required servo force. The system consists of the following components (Figure 7):
– Steering column that connects the steering pinion with the steering wheel inside the vehicle.
– Steering pinion that converts the rotating steering movement into the linear movement of the rack.

Figure 7: Variants of electromechanical power-assisted steering
a) Servo unit on the steering column,
b) Servo unit on a second pinion,
c) Paraxial servo unit.
1 Toothing for steering wheel, 2 Upper steering column, 3 Torque sensor, 4 Electric motor, 5 ECU, 6 Helical-gear drive, 7 Intermediate steering shaft, 8 Steering-spindle connection, 9 Mechanical rack-and-pinion steering, 10 Steering box, 11 Gaiter, 12 Tie rod, 13 Outer link, 14 Drive pinion (second pinion), 15 Recirculating-ball and toothed-belt gear.



– Rack connected to the wheels via the tie rods and links.
– Sensors to record the information required to calculate the necessary supporting steering torque.
– Servo unit, consisting of an electric ECU with microprocessor, electric motor and reduction gear which generates the assisting steering torque and links into the steering train.

Variants
The mechanical coupling of the motor to the steering box can be set up as a steering-column, steering-pinion or rack drive.

Servo unit on the steering column
The servo unit is integrated along with its electronics in the steering column (Figure 7a). It is connected via the intermediate steering shaft with universal joints to the mechanical rack-and-pinion steering. The torque generated by the electric motor is converted via the helical-gear drive (Figure 8) into an assisting torque and transmitted to the steering column. The sensor technology and the torsion-bar spring are located next to the helical-gear drive.

This system is used in vehicles with low steering forces (micro and compact cars).

Servo unit on a second pinion
The servo unit is mounted on a second pinion (Figure 7b), meaning that the sensor unit and drive unit can be separate. The helical-gear drive (Figure 8) converts the torque provided by the electric motor into the servo assisting torque and transmits it to the rack. The drive-pinion ratio is not dependent on the steering ratio; this makes possible a power-optimized design with 10 % to 15 % more system power.

This system is used in mid-size vehicles.

Paraxial servo unit
To convert the rotary movement of the steering wheel into a linear motion of the rack, this system uses a gear concept consisting of toothed belt and recirculating-ball gear. A system with ball recirculation is used in the ball-and-screw spindle drive. The ball chain is returned through a channel integrated in the recirculating-ball nut. The slip-free toothed belt is able to transmit the torque reliably.

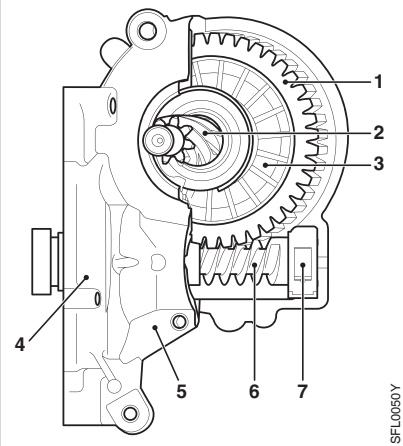
This system is used in vehicles with high steering forces (sports cars, upper mid-size, off-road vehicles, light commercial vehicles).

Servomotor
Brushed or brushless DC motors are used as servomotors. Depending on the required performance capability of the steering the design of the reduction gear, the torque generated by these motors is 3 to over 10 Nm.

Activation of the motor and hence the power consumption itself are demand-oriented; in other words, only when assisting power is actually required is the vehicle electrical system subjected to load extending beyond the stand-by power input.

Reduction gear
Primarily two principles are used in the reduction gear.

Figure 8: Helical-gear drive
1 Helical gear,
2 Drive pinion,
3 Overload safety,
4 Locating bearing,
5 Housing,
6 Worm,
7 Spring-damper element.

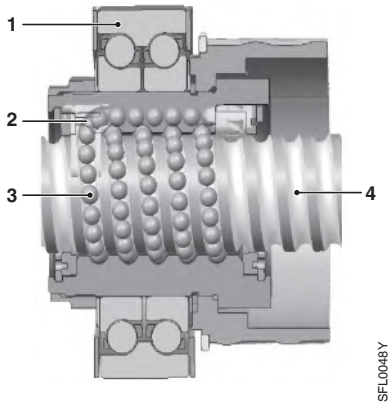


Helical-gear drive
When a helical-gear drive is used (Figure 8), the motor drives a worm which engages a helical gear and thereby acts on the steering output shaft, the steering pinion or a separate drive pinion (Figure 8). It is important to ensure in the design of the helical-gear drive that there is no tendency to self-lock when turning back as a result of the rack force.

Recirculating-ball gear
When a recirculating-ball gear is used (Figure 9), the motor drives, either directly or additionally reduced by an upstream belt-transmission stage, a ball-and-screw spindle drive, which converts the rotary movement into linear motion of the rack.

Torque sensor
To deliver the assisting force in line with demand, the driver manual torque is measured by the torque sensor integrated in the steering system, the required servo force is calculated in the electric ECU, and the motor is activated accordingly. The assisting force is delivered by the motor on the basis of torque control (see Torque sensor).

Figure 9: Recirculating-ball gear
1 Ball bearing,
2 Return channel,
3 Ball chain,
4 Rack.



Functions
Further functions are executed in the ECU software to optimize the steering behavior. The ECU evaluates further signals from the vehicle or the steering (e.g., driving speed, steering angle, steering torque, and steering speed) for this purpose. Thus, for example, the resetting movement of the steering wheel is optimized as a function of the steering angle and the driving speed by an additional motor torque. Likewise, a damping function is implemented with the information on driving speed and steering speed; this function supports the straight-ahead stability of the steering and hence of the vehicle at high driving speeds.

By networking the steering ECU with other ECUs in the vehicle network, it is possible to realize assistance functions with the electromechanical power-assisted steering to enhance comfort and safety.

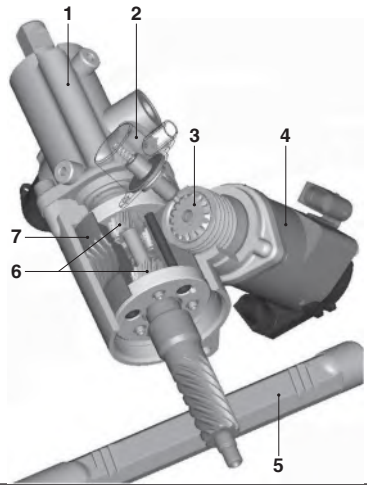
Demand-oriented control of the electric motor leads to considerable fuel savings of 0.3 *l*/100 km on average compared to hydraulic power steering with a pump driven by the vehicle engine. In urban driving, the fuel savings increase to 0.7 *l*/100 km.

In the event of a failure of the energy supply or steering assistance, the driver can continue to steer mechanically, but with higher manual steering torque.

Superimposed steering
With a superimposed-steering system, an additional steering angle can be added to or subtracted from the steering-wheel angle set by the driver at the steering wheel. This system is usually combined with a parameterizable hydraulic or an electric power-assisted steering system. Superimposed does not facilitate autonomous driving, but it does provide for a steering characteristic optimally adapted to the driving situation, and thus maximum comfort and directional stability. When networked with driving-dynamics control systems, it can further increase safety in critical driving situations by means of driver-independent steering adjustments. Such steering systems are already in series production, known as Active Steering at BMW and Dynamic Steering at Audi.

Technical solution
The angle superimposition independent of the driver's steering angle is currently effected by two different technical solutions.

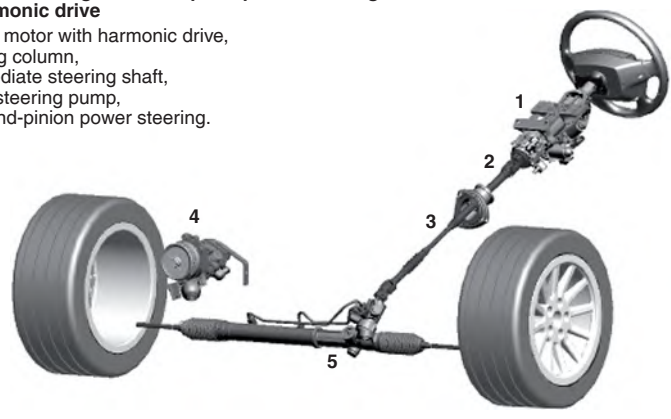
Figure 10: Planetary-gear set of superimposed steering
1 Valve,
2 Electromagnetic lock,
3 Worm,
4 Electric motor,
5 Rack,
6 Planetary gear,
7 Worm gear.



Planetary-gear set
A twin planetary-gear set with different gear ratios of the gear stages is integrated into a common planetary-gear carrier in the steering train (Figure 10). This means there is always a mechanical link between the steering wheel and the steered wheels. The different gear ratios of the gear stages mean that turning the planetary-gear carrier sets the additional steering angle. The angle setting is made by a electric motor that drives the worm gear of the planetary-gear carrier.

Harmonic drive
The steering-angle superimposition unit consists in this case of a harmonic drive and an electric motor with a hollow shaft (Figure 11). The highly compact design enables it to be integrated into the steering column without compromising the requirements with regard to installation space and crash behavior. The shaft at the steering-wheel end (input shaft) is positively connected (Figure 12) to a flex spline. The rotary movement of the steering wheel is transmitted via the toothing of the flex spline (teeth in the region of the flexible ball bearing point outwards) to the toothing of the internal gear (circular spline, teeth point inwards in the direction of the axle) for the output shaft. An elliptical inner rotor (shaft generator) located in the flex spline, which is driven by the electric motor, generates the superimposed steering angle via the different number

Figure 11: Steering train of superimposed steering with harmonic drive
1 Electric motor with harmonic drive,
2 Steering column,
3 Intermediate steering shaft,
4 Power-steering pump,
5 Rack-and-pinion power steering.



of teeth between the flex spline and the circular spline. The races of the flexible ball bearing follow the shaft motion of the elliptical inner rotor and transmit this shaft motion to the flex spline. Here, too, there is always a mechanical link between the steering wheel and the steered wheels via the toothing of the harmonic drive.

In the passive state, the electric motor is blocked by an electromechanical lock, thereby ensuring direct mechanical through-drive for the steering movement.

Activation concept
The superimposed-steering ECU checks the plausibility of the required sensor information and evaluates this information. It calculates the setpoint angle for the electric motor and generates via an integrated driver stage the pulse width-modulated signals for activating the electric motor. This is implemented as a brushless DC motor with integrated rotor position sensor. The maximum motor current is 40 A at a vehicle system voltage of 12 V. The rotor position sensor enables the control unit to control the electronic commutation and thus the direction of rotation of the motor. It also calculates and checks the total set additional steering angle using a summation algorithm in the control unit software.

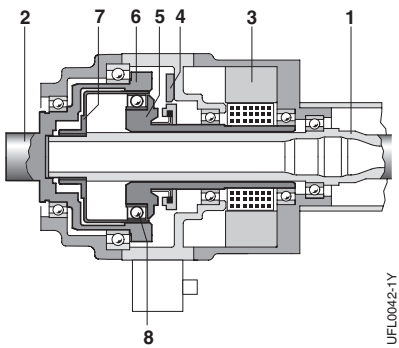
The effective steering angle, the sum total of the steering-wheel angle and the superimposed angle of the electric motor, is calculated by the ECU, and made available on the vehicle communication bus to the partner ECUs.

Setpoint value
The setpoint value for the effective steering angle formed in the superimposed-steering ECU is made up of the partial setpoint value for steering comfort and the partial setpoint value for vehicle stabilization. The signals required to calculate these variables are read in by the control unit via the CAN bus.

The partial setpoint value for the steering comfort is implemented as a speed-dependent, variable steering ratio. The value is calculated from the input variables of steering angle and driving speed. When the vehicle is stationary and at low driving speeds, an angle is added to the steering angle set by the driver. This makes the steering ratio more direct. The driver can turn the wheels fully with less than one complete steering wheel revolution. This steering-angle addition is continuously reduced as driving speed increases. From speeds of roughly 80 – 90 km/h, a proportion is subtracted from the driver's steering angle and steering becomes more indirect. This ensures straight-running vehicle stability at high speeds and at the same time prevents the driver from losing control of the vehicle due to excessively fast steering movements.

For the calculation of the partial setpoint value for vehicle stabilization – in addition to the steering angle and the driving speed – the vehicle movement is measured using the sensors for the vehicle yaw rate and lateral acceleration. The superimposed steering uses the driving-dynamics control sensors for this purpose. In the same way as the driving-dynamics control, a calculation model that runs in the ECU software calculates the reference vehicle movement. In the event of a deviation of the actual vehicle movement from the reference movement, the steering is activated to stabilize the vehicle. The two systems continuously

Figure 12: Actuator of superimposed steering with harmonic drive
1 Input shaft,
2 Output shaft,
3 Electric motor,
4 Rotor-position sensor,
5 Elliptical inner rotor (shaft generator),
6 Internal gear (circular spline),
7 Flex spline,
8 Flexible ball bearing.



exchange information so that the controllers of the driving-dynamics control and the superimposed steering cooperate to optimal effect.

Safety concept
All of the internal and external signals that are used are continuously monitored in the control unit and their plausibility checked. If a sensor signal is no longer plausible, the additional steering function on which it is based is first disabled. For example, if the yaw-rate sensor that measures the rotation of the vehicle around its vertical axis (yaw rate) fails, the yaw-rate control of the superimposed-steering system is disabled. The variable steering ratio remains active.

If safe activation of the electric motor is no longer possible due to a fault, the system is completely shut down and direct steering-wheel through-drive is ensured by self-locking of the gear stage and by an electromechanical lock. This fallback level is always automatically active when the internal-combustion engine is not running or even if there is no electrical supply voltage.

Power-assisted steering systems for commercial vehicles

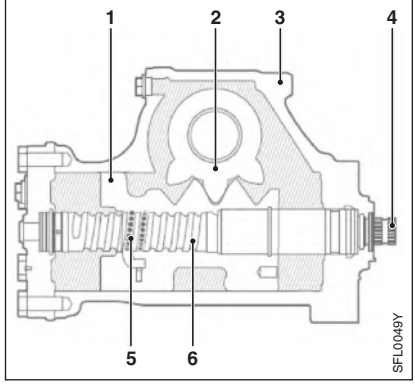
Power-assisted steering with all-hydraulic transmission

Hydrostatic steering systems are hydraulic power-assisted steering systems. The steering force of the driver is hydraulically boosted and transmitted exclusively by hydraulic means to the steered wheels. Because there is no mechanical connection, the maximum permissible speed is limited by national regulations. In Germany, this is 25 km/h. Depending on the system configuration and emergency steering properties, approval up to a speed of 62 km/h is possible. Use of these systems is therefore confined to machinery and special vehicles.

Single-circuit power-assisted steering system for commercial vehicles

Commercial vehicles are usually equipped with ball-and-nut power steering (Figure 13). The control valve is integrated into the steering box and together with the steering worm forms a single unit. The rotary movement of the steering wheel is transmitted via an endless ball chain to

Figure 13: Ball-and-nut power steering
1 Piston,
2 Steering-sector shaft,
3 Housing,
4 Steering-spindle connection,
5 Ball nut with ball chain,
6 Steering worm.

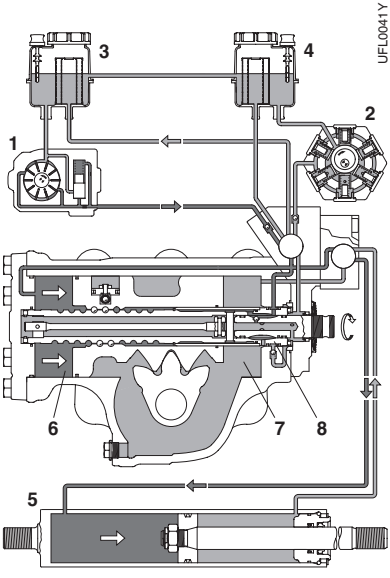


the ball-and-nut. Short toothing on the ball-and-nut meshes with the toothing on the segment shaft. The generated rotary movement of the segment shaft is transmitted via a steering arm to the steering linkage and the steered wheels.

The servo force is applied, as with rack-and-pinion power steering, by a rotary-slide valve. The working cylinder is formed by a sealing surface between the ball-and-nut housing and the steering box. Because no additional lines are required outside the housing, a robust and compact steering box with high power output is created.

Figure 14: Dual-circuit power-assisted steering

1 Steering pump 1,
2 Steering pump 2,
3 Oil reservoir 1,
4 Oil reservoir 2,
5 Working cylinder,
6 Left cylinder chamber,
7 Right cylinder chamber,
8 Dual-circuit steering valve.



Dual-circuit power-assisted steering system for heavy-duty commercial vehicles

Dual-circuit steering systems (Figure 14) are required when the operating forces needed at the steering wheel exceed the statutory requirements in ECE-R79 [1] if the power-assistance system fails. These steering systems feature hydraulic redundancy. Both steering circuits in these systems are functionally tested by means of flow indicators and a fault is indicated to the driver. The pumps for supplying the independent steering circuits must be driven in various ways (e.g., engine-dependent, driving speed-dependent, or electrically). If a circuit fails, caused for example by a fault in the steering system or engine failure, the vehicle can be steered with the still operational redundant circuit in accordance with the statutory requirements.

Dual-circuit systems take the form of ball-and-nut power steering with an integrated second steering valve. This second valve controls an additionally externally installed working cylinder and thereby provides the redundancy in relation to the existing servo system in the ball-and-nut steering.

References

[1] ECE-R79: Uniform provisions concerning the approval of vehicles with regard to steering equipment.

Brake systems

Definitions and principles

(based on ISO 611 [1] and DIN 70024 [2]).

Brake equipment
All the vehicle brake systems whose functions are to reduce vehicle speed or bring the vehicle to a halt, or to hold the vehicle stationary if already halted.

Brake systems
Service brake system
All the elements, the action of which may be regulated, allowing the driver to reduce, directly or indirectly, the speed of a vehicle during normal driving or to bring the vehicle to a halt.

Secondary brake system
All the elements, the action of which may be regulated, allowing the driver to reduce, directly or indirectly, the speed of a vehicle or to bring the vehicle to a halt in case of failure of the service brake system.

Parking brake system
All the elements allowing the vehicle to be held stationary mechanically even on an inclined surface, and particularly in the absence of the driver.

Continuous-operation brake system
System of components which allows the driver to reduce the vehicle's speed or descend a long downhill gradient at a virtually constant speed with practically no wear to the friction brakes. A continuous-operation brake system may incorporate one or more retarders.

Automatic brake system
All the elements which automatically brake the trailer as a result of intended or accidental separation from the tractor vehicle.

Electronic brake system (ELB, EHB)
Brake system controlled by an electrical signal generated and processed by the control transmission system. An electrical output signal controls components which generate the application force.

Component parts
Energy-supplying device
Parts of a brake system which supply, regulate and, if necessary, condition the energy required for braking. It terminates at the point where the transmission device starts, where the various circuits of the brake systems, including the circuits of accessories if fitted, are isolated either from the energy supplying device or from each other.

The energy source is that part of an energy-supplying device which generates the energy. It may be located remotely from the vehicle (e.g. in the case of a compressed-air braking system for a trailer) or may be the muscular force of the driver.

Control device
Parts of a brake system which initiate the operation and control the effect of this brake system. The control signal can be conveyed within the control device by, for example, mechanical, pneumatic, hydraulic or electrical means, including the use of auxiliary energy or non-muscular force.

The control device is defined as starting at the component to which the control force is directly applied. It can be operated:
– By direct application of force by the driver by hand or foot,
– by indirect action of the driver or without any action (only in the case of trailers),
– by varying the pressure in a connecting line, or the electric current in a cable, between the tractor vehicle and the trailer at the time when one of the brake systems on the tractor vehicle is operated, or if it fails,
– by the inertia of the vehicle or by its weight or of one of its main component parts.

The control device is defined as ending at the point at which the braking energy is distributed, or where part of the energy is diverted to control braking energy.

Transmission device
Parts of a brake system which transmit the energy distributed by the control device. It starts either at the point where the control device terminates or at the point where the energy supplying device terminates. It terminates at those parts of the brake system in which the forces opposing the vehicle's movement, or its tendency towards movement, are generated. It can, for example, be mechanical, hydraulic, pneumatic (pressure above or below atmospheric), electric, or combined (for example hydromechanical, hydropneumatic).

Brake
Parts of a brake system in which the forces opposing the vehicle's movement, or its tendency towards movement, are developed, such as friction brakes (disk or drum) or retarders (hydrodynamic or electrodynamic retarders, exhaust brakes).

Auxiliary device of the tractor vehicle for a trailer
Parts of a brake system on a tractor vehicle which are intended to supply energy to, and control, the brake systems on the trailer. It comprises the components between the energy supplying device of the tractor vehicle and the supply-line coupling head (inclusive), and between the transmission device(s) of the tractor vehicle and the control-line coupling head (inclusive).

Brake-system types relating to the energy supplying device
Muscular-energy brake system
Brake system in which the energy necessary to produce the braking force is supplied solely by the physical effort of the driver.

Energy-assisted brake system
Brake system in which the energy necessary to produce the braking force is supplied by the physical effort of the driver and one or more energy supplying devices.

Non-muscular-energy brake system
Brake system in which the energy necessary to produce the braking force is supplied by one or more energy-supplying devices excluding the physical effort of the driver. This is used only to control the system.

Note: A brake system in which the driver can increase the braking force, in a state of totally failed energy, by muscular effort acting on the system, is not included in the above definition.

Inertia brake system
Brake system in which the energy necessary to produce the braking force arises from the approach of the trailer to its tractor vehicle.

Gravity brake system
Brake system in which the energy necessary to produce the braking force is supplied by the lowering of a component part of the trailer (e.g. trailer drawbar) due to gravity.

Definitions of brake systems relating to the arrangement of the transmission device
Single-circuit brake system
Brake system having a transmission device embodying a single circuit. The transmission device comprises a single circuit if, in the event of a failure in the transmission device, no energy for the production of the application force can be transmitted by this transmission device.

Multi-circuit brake system
Brake system having a transmission device embodying several circuits. The transmission device comprises several circuits if, in the event of a failure in the transmission device, energy for the production of the application force can still be transmitted, wholly or partly, by this transmission device.

Definitions of brake systems relating to vehicle combinations

Single-line brake system
Assembly in which the brake systems of the individual vehicles act in such a way that the single line is used both for the energy supply to, and for the control of, the brake system of the trailer.

Dual- or multi-line brake systems
Assembly in which the brake systems of the individual vehicles act in such a way that several lines are used separately and simultaneously for the energy supply to, and for the control of, the brake system of the trailer.

Continuous brake system
Combination of brake systems for vehicles forming a road train. Characteristics:
– From the driving seat, the driver can operate a directly operated control device on the tractor vehicle and an indirectly operated control device on the trailer by a single operation and with a variable degree of force.
– The energy used for the braking of each of the vehicles forming the combination is supplied by the same energy source (which may be the muscular effort of the driver).
– Simultaneous or suitably phased braking of the individual units of a road train.

Semi-continuous brake system
Combination of brake systems for vehicles forming a road train. Characteristics:
– The driver, from his driving seat, can gradually operate a directly operated control device on the tractor vehicle and an indirectly operated control device on the trailer by a single operation.
– The energy used for the braking of each of the vehicles forming the road train is supplied by at least two different energy sources (one of which may be the muscular effort of the driver).
– Simultaneous or suitably phased braking of the individual units of a road train.

Non-continuous brake system
Combinations of the brake systems of the vehicles forming a road train which is neither continuous nor semi-continuous.

Brake-system control lines
Wiring and conductors: These are employed to conduct electrical energy.

Tubular lines: Rigid, semi-rigid or flexible tubes or pipes used to transfer hydraulic or pneumatic energy.

Lines connecting the brake equipment of vehicles in a road train
Supply line: A supply line is a special feed line transmitting energy from the tractor vehicle to the energy accumulator of the trailer.

Brake line: A control line is a special control line by which the energy essential for control is transmitted from the tractor vehicle to the trailer.

Common brake and supply line:
Line serving equally as brake line and as supply line (single-line brake system).

Secondary-brake line: Special actuating line transmitting the energy from the tractor vehicle to the trailer essential for secondary braking of the trailer.

Braking mechanics
Mechanical phenomena occurring between the start of actuation of the control device and the end of the braking action.

Gradual braking
Braking which, within the normal range of operation of the control device, permits the driver, at any moment, to increase or reduce, to a sufficiently fine degree, the braking force by operating the control device. When an increase in braking force is obtained by the increased action of the control device, an inverse action must lead to a reduction in that force.

Brake-system hysteresis: Difference in control forces between application and release at the same braking torque.

Brake hysteresis: Difference in application force between application and release at the same braking torque.

Forces and torques
Control force F_c: Force exerted on the control device.

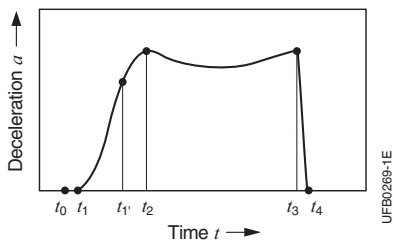
Application force F_s: On friction brakes, the total force applied to a brake lining and which causes the braking force by the effect of friction.

Braking torque: Product of frictional forces resulting from the application force and the distance between the points of application of these forces and the axis of rotation of the wheels.

Total braking force F_f: Sum of the braking forces at the tire contact patches of all the wheels and the ground, produced by the effect of the brake system, and which oppose the movement or the tendency of the vehicle to move.

Braking-force distribution: Specification of braking force according to axle, given in % of the total braking force F_f. Example: front axle 60 %, rear axle 40 %.

Figure 1: Times and deceleration during braking to a stop



before t_0: Reaction time,
t_0: Initial application of force on control device,
t_1: Start of deceleration,
t_1': End of pressure build-up time,
t_2: Fully developed deceleration,
t_3: End of maximum retardation,
t_4: End of braking operation (vehicle stationary),
$t_1 - t_0$: Initial response time,
$t_1' - t_1$: Pressure build-up time,
$t_3 - t_2$: "Mean maximum retardation" range,
$t_4 - t_1$: Active braking time,
$t_4 - t_0$: Total braking time.

Brake coefficient C^*: Defines the relationship between the total peripheral force of a given brake and the brake's application force.

$$C^* = \frac{F_u}{F_s} .$$

F_u Total peripheral force,
F_s Application force.
The mean is employed when there are variations in application forces at individual brake shoes (i number of brake shoes):

$$F_s = \Sigma \frac{F_{si}}{i} .$$

Time periods
Reaction time (see Figure 1): The time that elapses between perception of the state or object which induces the response, and the point at which the control device is actuated (t_0).

Actuating time of the control device: Elapsed time between the moment when the component of the control device (t_0) on which the control force acts starts to move, and the moment when it reaches its final position corresponding to the applied control force (or its travel). (This is equally true for application and release of the brakes).

Initial response time $t_1 - t_0$: Elapsed time between the moment when the component of the control device on which the control force acts starts to move and the moment when the braking force takes effect.

Pressure build-up time $t_1' - t_1$: Period that elapses between the point at which the braking force starts to take effect and the point at which a certain level is reached (75 % of asymptotic pressure in the wheel-brake cylinder as per EU Directive 71/320/ EEC [3], Annex III/2.4).

Initial response and pressure build-up time: The sum of the initial response and pressure build-up times is used to assess how the brake system behaves over time until the moment at which full braking effect is reached.

Active braking time $t_4 - t_1$: Elapsed time between the moment when the braking force starts to take effect and the moment when the braking force ceases. If the

vehicle stops before the braking force ceases, the time when motion ceases is the end of the active braking time.
Release time: Elapsed time between the moment when the control device starts to release and the moment when the braking force ceases.

Total braking time $t_4 - t_0$: Elapsed time between the moment when the control device on which the control force acts starts to move and the moment when the braking force ceases. If the vehicle stops before the braking force ceases, the time when motion ceases is the end of the active braking time.

Braking distance s
Distance traveled by the vehicle during the total braking time. If the time when motion ceases constitutes the end of the total braking time, this distance is called the "stopping distance".

Braking work W
Integral of the product of the instantaneous braking force F_f and the elementary movement ds over the braking distance s:

$$W = \int_0^s F_f \, ds \, .$$

Instantaneous braking power P
Product of the instantaneous total braking force F_f and the vehicle's road speed v:

$$P = F_f \, v \, .$$

Braking deceleration
Reduction of speed obtained by the brake system within the considered time t. A distinction is made between the following:

Instantaneous braking deceleration

$$a = \frac{dv}{dt} \, .$$

Mean braking deceleration over a period of time
The mean braking deceleration between two points in time t_B and t_E is

$$a_{mt} = \frac{1}{t_E - t_B} \int_{t_B}^{t_E} a(t) \, dt;$$

This means that:

$$a_{mt} = \frac{v_E - v_B}{t_E - t_B} \, ,$$

where v_B and v_E are the vehicle speeds at the times t_B and t_E.

Mean braking deceleration over a specific distance
The mean braking deceleration over the distance between two points s_B and s_E is:

$$a_{ms} = \frac{1}{s_E - s_B} \int_{s_B}^{s_E} a(s) \, ds \; ;$$

This means that:

$$a_{ms} = \frac{v_E^2 - v_B^2}{2 \, (s_E - s_B)}$$

where v_B and v_E are the vehicle speeds up to the points s_B and s_E.

Mean braking deceleration over the total braking distance
The mean braking deceleration is calculated according to the equation:

$$a_{ms0} = \frac{-v_0^2}{2 \, s_0} \, ,$$

where v_0 relates to the time t_0 (special instance of a_{ms} where $s_E = s_0$).

Mean fully developed deceleration d_m
Mean fully developed deceleration over the distance determined by the conditions $v_B = 0.8\, v_0$ and $v_E = 0.1\, v_0$ thus:

$$d_m = \frac{v_B^2 - v_E^2}{2 \, (s_E - s_B)}$$

The mean fully developed deceleration is used in ECE Regulation 13 [6] as a measure of the effectiveness of a brake system. Since positive values for d_m are used here, the mathematic sign has been reversed in this case. (In order to establish a relationship between braking distance and braking deceleration, braking deceleration must be expressed as a function of the distance traveled.)

Braking factor z
Ratio between the total braking force, F_f, and the permissible total static weight, G_s, exerted on the axle or axles of the vehicle:

$$z = \frac{F_f}{G_s} \, .$$

Legal regulations

General Certification for a vehicle with regard to its brake system may only be granted when the brake system complies with the following regulations:
– § 41 StVZO [4] (German Road Vehicle Registration Regulation) in conjunction with § 72 StVZO [5] and the associated directives.
– Council Directive of the European Community (RREG) 71/320/EEC [3], associated Amending Directives and Annexes.
– ECE Regulations R13 [6], R13H [7], R78 [8].

In § 41 StVZO, the requirements placed on the brake system differ depending on the type, gross weight rating, application, date of registration, and type-determined top speed. In the EC Directives, the requirements are allocated to individual vehicle categories. The vehicle categories are as follows:
– M1, M2, M3: Passenger cars with at least four wheels,
– N1, N2, N3: Commercial vehicles with at least four wheels,
– O1, O2, O3, O4: Trailers and semitrailers,
– L1, L2, L3, L4: Motorcycles, threewheelers.

The values stipulated in § 41 StVZO with regard to mean fully developed deceleration do not apply, for example, to the recurring inspections required in Germany for registered vehicles on the road (general inspection, safety inspection). In these inspections, the requirements of § 29 StVZO [9], s. 1, Annex VIII in conjunction with Annex VIIIa, Guideline for performing the general inspection and Guideline for performing the safety inspection, apply.

The requirements in § 41 StVZO and ECE-R13H with regard to brake equipment are essentially identical. However, ECE Regulations R13, R13H and R78 have been further updated and also contain, for example, regulations covering electronically controlled brake systems. The prescribed braking effects must be determined in accordance with Directive 71/320/EEC, s. 1.1.2, Annex II, amended by Directive 98/12/EC [10], or in accordance with § 41 StVZO, s. 12.

Requirements placed on brake systems
(as per § 41 StVZO, EC Directive 71/320/EEC, ECE-R13, as at 2009)
Category M and N vehicles must comply with the provisions which pertain to the service, secondary and parking brake systems. The brake systems may have common components. Such vehicles must have at least two mutually independent control devices for the brake systems, one of which must be lockable. The control devices must be fitted with separate transmission devices, each must be able to continue operating when the other fails. The distribution of braking force between the individual axles is prescribed and must be sensible. If a malfunction occurs, it must be possible to achieve the prescribed secondary braking effect with the remaining operational part of the brake system or with the vehicle's other brake system without the vehicle departing from its lane.

Vehicles in categories M2 and N2 and above must be fitted with automatic antilock braking facilities. Regulation EC 661/2009 [11] stipulates that from 1 November 2011 all new vehicle models and from 2014 all new category M1 and N1 vehicles to be brought onto the road must be equipped with an electronic driving-dynamics control system (Electronic Stability Program). This regulation also applies, with the exception of off-road vehicles as per Directive 2007/46/EC [12], Annex II, Part A, to vehicles in the following categories:
– M2 and M3, except vehicles with more than three axles, articulated buses and buses of category 1 or A,
– N2 and N3, except vehicles with more than three axles, tractor units with a total weight of between 3.5 and 7.5 t and special-purpose vehicles as per Directive 2007/46/EC, Annex II, Part A,
– O3 and O4, with pneumatic suspension, except vehicles with more than three axles, trailers for heavy transports and trailers with areas for standing passengers.

The implementation of this regulation for the vehicle categories, except categories M1 and N1, is defined in Regulation EC 661/2009 [11], Annex V.

Continuous-operation brake systems
Continuous-operation brake systems are additionally used to relieve the strain on the service brake on long downhill gradients. Category M3 vehicles for local and long-distance duty (buses weighing more than 5.5 t (except city buses)) and other category N2 and N3 vehicles with a gross weight rating of more than 9 t (Directive 71/320/EEC, §41 StVZO, s. 15) must be equipped with a retarder. Exhaust brakes or similar facilities are classed as retarders. The retarder must be designed to hold the vehicle when fully laden when driving on a downhill gradient of 7% and a distance of 6 km at a speed of 30 km/h.

Category O trailers
Category O1 trailers are not required to have their own brake system; a securing connection to the tractor vehicle is sufficient. Category O2 trailers and above must be fitted with service and parking brake systems, which may have common components. The distribution of braking force between the individual axles is prescribed in Directive 71/320/EEC. It must be sensibly distributed between the axles.

Category O3 trailers and above (ECE), as well as trailers and semitrailers with a gross weight rating of more than 3.5 t and a type-determined top speed of more than 60 km/h (§41b StVZO, s. 2), must be equipped with an antilock braking facility. Semitrailers only need to be equipped with ABS if the gross weight rating reduced by the fifth-wheel load exceeds 3.5 t.

Category O3 trailers (existing types) which are registered for use on public roads after 11.07.2014 or 01.11.2014 must be equipped with an electronic driving-dynamics control system (Electronic Stability Program). For new types, this regulation will already apply from 01.11.2011 or 11.07.2012 (Regulation EC 661/2009 [11], Annex V).

Category O2 trailers and below must be equipped with inertia brake systems. The trailer must brake automatically if it becomes decoupled from the tractor vehicle while moving, or (for trailers weighing less than 1.5 t) it must be equipped with a securing connection to the tractor vehicle.

Category L vehicles
Motorized two- and three-wheeled vehicles must be equipped with 2 mutually independent brake systems. In the case of duty category L5 three-wheel vehicles, the two brake systems must both act on all the wheels. A parking brake system must be fitted.

Tractor vehicles and trailers with compressed-air brake systems
The compressed-air connections between the individual vehicles must be of the dual- or multi-line design. This ensures that the compressed-air brake system in the trailer can also be refilled during the braking operation. When the service brake system on the tractor unit is operated, the service brake system on the trailer must also be operated with a variable degree of force. If a fault occurs in the service brake system of the tractor unit, that part of the system not affected by the fault must be capable of braking (controlling) the trailer with a variable degree of force. If one of the connecting lines between tractor unit and trailer is interrupted or develops a leak, it must still be possible to brake the trailer, or it must brake automatically.

The energy accumulators of the service brake systems must be designed such that, after eight full operations of the service brake, at least the requested secondary braking effect is furnished on the ninth braking. The energy accumulators must not be replenished during this test. The braking effect of the individual vehicles is prescribed as a function of the pressure at the “Brake” coupling head in Directive 71/320/EEC.

Vehicles with antilock braking facilities
Antilock braking facilities must comply with Directives 71/320/EEC, Annex X and ECE-R13, Annex 13. An antilock braking facility is part of a service brake system which automatically controls the slip in the direction of wheel rotation at one or more

of the vehicle's wheels during the braking operation. The requirements placed on the antilock braking facility differ, depending on the category, for vehicles ABS categories 1, 2 and 3, for trailers ABS categories A and B.

Essential requirements placed on the antilock braking facility (Category 1) are:

- Locking of the directly controlled wheels under braking must be prevented at speeds of over 15 km/h on all road surfaces.
- Directional stability and maneuverability must be maintained. Under a μ-split condition (extremely different coefficients of friction between the left and right wheels), steering corrections of 120° during the first two seconds and 240° in total are permissible.
- There must be a special visual warning system (yellow warning signal) to indicate electrical faults.
- Motor vehicles (except Categories M1 and N1) with ABS which are equipped to tow a trailer with ABS must be fitted with a separate visual warning system (yellow warning signal) for the trailer. Transfer must be effected via pin 5 of the electrical plug-in connection as per ISO 7638 [13].
- The energy accumulators of the service brake system in vehicles with ABS must be designed such that the prescribed secondary braking effect is still achieved even after a controlled braking operation of longer duration ($t = v_{max}/7$, at least 15 seconds) and then four uncontrolled full-braking operations without energy replenishment.

Requirements and test conditions

The required values and test conditions must be applied during the test in accordance with § 19 [14] and § 20 StVZO [15]. Departures from the test method described in Directive 71/320/EEC, Annex II, s. 1.1.2, last amended by Directive 98/12/EC and the test method specified in § 41 StVZO, s. 12, are permitted – especially during verification checks as per § 29 StVZO – if the condition and the effect can be ascertained by other means (§ 41 StVZO, s. 12). When vehicles to be newly registered are tested, a higher braking deceleration corresponding to the usual easing of the braking effect must be achieved; furthermore, sufficient state-of-the-art continuous brake duty must be guaranteed for longer stretches of downhill driving.

Minimum retardation and max. permissible control forces during the general inspection as per § 29 StVZO

The required retardation values are determined on test benches and only in exceptional cases in road tests. The required values are maximum values, because the time response which is needed to determine the mean fully developed deceleration is not measured in these recurring inspections.

Structure and organization of brake systems

Essential requirements
The brake systems in motor vehicles must comply with the requirements of Directive 71/320/EEC, ECE-R13 Part 1, ECE-R13 Part 2 and ECE-R13 H, and other country-specific regulations. Motor vehicles must be equipped with two separate brake systems, one of which must be lockable. The brake systems must have separate control devices. In the event of a fault in the service brake system, it must still be possible for at least two wheels (not on the same side) to be braked.

Types of brake system
The brake systems comprise the service brake and parking brake systems and (in commercial vehicles and motor buses) the continuous-operation (retarder) brake system. The requested secondary brake system normally comes into play when a fault occurs in the service brake system. Special-purpose vehicles with special requirements may also have special braking functions such as a hill-climbing brake or an anti-jackknifing brake.

Type of force generation
When it comes to how the force is generated, there are three different types of system: muscular-energy, energy-assisted, and non-muscular-energy brake systems. In muscular-energy systems, the muscular force of the driver alone is effective, in energy-assisted systems this is boosted by booster systems (brake booster), and in non-muscular-energy systems, the driver's control force acts only as a control variable. The maximum required control forces are prescribed for each type of vehicle.

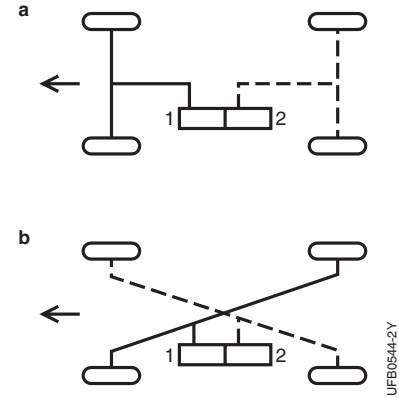
Transmission device
Force is transmitted from the control device to the wheel brakes by mechanical, hydraulic, pneumatic or electric means. Mechanical force transmission is only customary and prescribed for parking brake systems (§41 StVZO, s. 5).

Force transmission for the service brake system is performed via two separate brake circuits by hydraulic or pneumatic means so that at least one brake circuit remains in operation in the event of a fault.

Electrical brake operation has up to now only been used in electrically acting parking brake systems (see Electromechanical parking brake systems).

Brake-circuit configuration
Configuration of the brake circuits is governed by DIN 74000 [16]. In category M1 vehicles (passenger cars), the brake-circuit configuration is often diagonal (Figure 2b). But this is only possible in conjunction with suitable front-axle geometry (steering offset negative or neutral). In all other vehicle categories, the II configuration is used (Figure 2a). Here the front axle forms one of the brake circuits and the rear axle forms the other. All other brake-circuit configurations as per DIN 74000 are rarely used today and are therefore no longer described in these pages. The direct demand for a dual-circuit design of the transmission device is prescribed in §41 StVZO, s. 16 only for motor buses.

Figure 2: Variants of brake-circuit configuration
a) II configuration,
b) X configuration.
1 Brake circuit 1,
2 Brake circuit 2.
← Direction of travel.



Braking-force distribution
Directives 71/320/EEC, ECE-R13 and ECE-R13H also place requirements on braking-force distribution between the individual axles. This must be sensibly distributed in all load states between the axles. Braking-force distribution can be effected on the one hand through an assembly-related configuration of the wheel brakes and on the other hand through a vehicle-related configuration. Among other things, the center-of-gravity height, the wheelbase and the empty-empty ratio of the vehicle are taken into consideration.

In commercial vehicles, according to the diagrams in Directive 71/320/EEC, braking-force distribution is also dependent on the pressure at the "Brake" coupling head. Vehicle-related configuration of braking-force distribution is effected by the integration of a braking-force limiter or an automatically acting braking-force metering device (automatic load-sensitive braking-force metering).

In modern vehicles braking-force distribution is integrated as an additional function into the electronic wheel-slip control system (antilock braking facility, driving-dynamics control).

Assemblies
Brake systems in motor vehicles consist of the following assemblies, which differ in design depending on whether the system is hydraulic or pneumatic: energy supply, control devices, transmission devices, control facilities, wheel brakes, and auxiliary devices.

References
[1] ISO 611: Road vehicles – Braking of automotive vehicles and their trailers – Vocabulary.
[2] DIN 70024: Vocabulary for components of motor vehicles and their trailers.
[3] EC Directive 71/320/EEC: Council Directive of 26 July 1971 on the approximation of the laws of the Member States relating to the braking devices of certain categories of motor vehicles and their trailers.
[4] §41 StVZO: Brakes and wheel chocks.
[5] §72 StVZO: Entry into force and transitional provisions.
[6] ECE-R13: Standard conditions for approval of category M, N and O vehicles with regard to brakes.
[7] ECE-R13H: Standard conditions for approval of passenger cars with regard to brakes. Day of entry into force: 11 May, 1998.
[8] ECE-R78: Standard conditions for approval of category L1, L2, L3, L4 and L5 vehicles with regard to brakes.
[9] §29 StVZO: Inspection of motor vehicles and trailers.
[10] EC Directive 98/12/EC of the Commission of 27 January 1998 adapting to technical progress Council Directive 71/320/EEC on the approximation of the laws of the Member States relating to the braking devices of certain categories of motor vehicles and their trailers.
[11] Regulation (EC) No. 661/2009 of the European Parliament and of the Council of 13 July 2009 concerning type-approval requirements for the general safety of motor vehicles, their trailers and systems, components and separate technical units intended therefore.
Published in the Official Journal of the European Union L200 of 31. July 2009.
[12] Directive 2007/46/EC of the European Parliament and of the Council of 5 September 2007 establishing a framework for the approval of motor vehicles and their trailers, and of systems, components and separate technical units for such vehicles (Framework Directive).
[13] ISO 7638: Road vehicles – Connectors for the electrical connection of towing and towed vehicles.
[14] §19 StVZO: Granting and effectiveness of design certification.
[15] §20 StVZO: General Certification for types.
[16] DIN 74000: Hydraulic braking systems; dual circuit brake systems; symbols for brake circuits diagrams.

Brake systems for passenger cars and light commercial vehicles

Subdivision of passenger-car brake systems

Brake systems for passenger cars and light utility vehicles must comply with the requirements of various directives and statutory provisions, e.g., 71/320/EEC [1], ECE R13 [2], ECE R13-H [3] and in Germany §41 StVZO [4]. The requirements with regard to functioning, effect and test methods are set out in these regulations.

The entire system is subdivided into the service-brake system, the parking-brake system and the secondary-brake system.

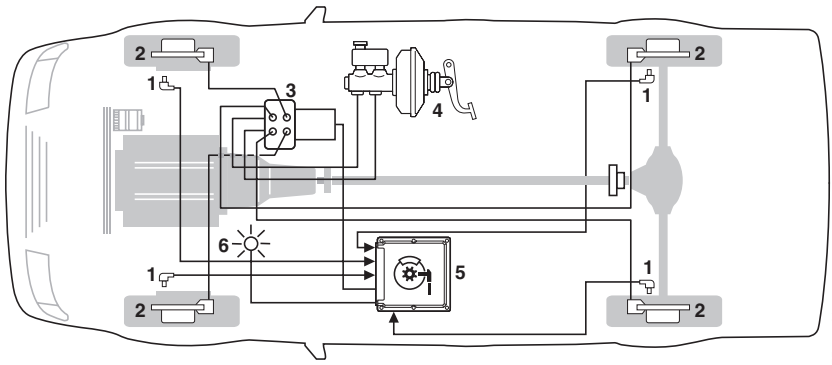
Service-brake system
The service-brake system allows the driver to reduce with graduable effect the speed of a vehicle during normal driving or to bring the vehicle to a halt. In passenger cars and light utility vehicles it is normally designed as an energy-assisted brake system.

The driver meters the braking effect steplessly by pressing on the brake pedal. Force is transmitted to the wheel brakes via the tandem brake master cylinder to two mutually independent hydraulic transmission devices (Figure 1). The service-brake system acts on all four wheels.

Parking-brake system
The parking-brake system ("handbrake") is an independent brake system which holds the vehicle stationary, even on a gradient and especially when the driver is not in the vehicle. The holding-stationary mechanism is integrated in the wheel brake. Legal requirements stipulate that the parking brake must have a consistently mechanical connection between the control device and the wheel brake, e.g., by means of a linkage or a control cable.

The parking brake is usually actuated by a handbrake lever next to the driver's seat, and in some cases also by a pedal. In the case of electrically actuated parking-brake systems, the parking brake is locked or released by means of an electrical operating device (switch). The

Figure 1: Hydraulic dual-circuit brake system
1 Wheel-speed sensors,
2 Wheel brakes (disk brakes, drum brakes also possible on the rear axle),
3 Hydraulic modulator (for antilock braking system and driving-dynamics control system),
4 Control device with brake booster, tandem brake master cylinder and expansion reservoir,
5 ECU (can be directly mounted on the hydraulic modulator),
6 Warning lamp for antilock braking system and driving-dynamics control.



service- and parking-brake systems are thus equipped with separate, individual control and transmission devices. The parking-brake system may be graduable in design, acting on the wheels of only one axle.

The holding-stationary effect is calculated according to ECE R13-H on a downhill gradient with fully laden vehicles. The downhill gradient for single vehicles is 20 %. If the vehicle is equipped to tow a trailer, the holding-stationary effect must also be achieved with an unbraked trailer on a downhill gradient of 12 %.

Secondary-brake system
In the event of a fault, e.g., leak or fractured pipe, it must still be possible with the operational part of the brake system to achieve at least the secondary braking effect – with the identical control force at the control device. It must be possible to meter the secondary braking effect, which in turn must be at least 50 % (ECE R13-H) or 44 % (§41 s. 4a). The vehicle must not leave its lane when the secondary brake is applied.

The secondary-brake system does not need to be an independent third brake system (in addition to the service- and parking-brake systems) with a special control device. Either the intact brake circuit of a dual-circuit service-brake system or a graduable parking-brake system can be used as the secondary-brake system.

Components of the passenger-car brake system

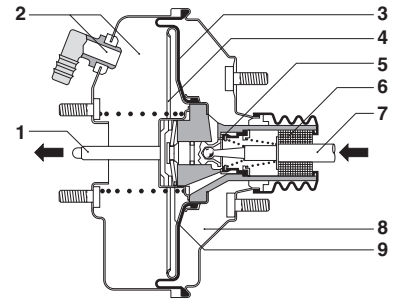
Control device
The control device comprises those parts of the brake system which initiate the effect of this brake system. When the service brake is applied, the force of the driver's foot acts on the brake pedal. The lever-transmitted pedal force is boosted in the brake booster, depending on its design, by a factor of 4 to 10 and acts on the piston in the brake master cylinder (Figure 1). The control force is converted into hydraulic pressure. Under full braking, this pressure ranges between 100 and 160 bar, depending on the system configuration.

Vacuum brake booster
Function
The brake booster reduces the applied control force required for the braking operation, but must not impair the sensitive graduation of the braking force and the feeling of the measure of braking.

Design
Brake boosters function as vacuum-operated boosters or also hydraulically. Hydraulic brake boosters are supplied by the power steering or by a separate hydraulic pump and pressure-accumulator devices.

Passenger-car brake systems are usually equipped with vacuum brake boosters. These vacuum boosters utilize the negative pressure generated in the intake manifold during the induction stroke on gasoline engines or the vacuum (0.5 – 0.9 bar) produced by a vacuum pump on

Figure 2: Vacuum brake booster
1 Push rod,
2 Vacuum chamber with vacuum connection,
3 Diaphragm,
4 Power piston,
5 Valve unit,
6 Air filter,
7 Piston rod,
8 Working chamber,
9 Reaction element.



diesel engines and on electric vehicles or hybrid vehicles to amplify the force applied by the driver's foot. A diaphragm separates the vacuum chamber with vacuum connection from the working chamber (Figure 2). The piston rod transmits the applied foot pressure to the working piston and the amplified force is passed to the brake master cylinder via the push rod.

Operating principle
If the brake is not operated, the vacuum chamber and working chamber are connected via the valve unit. Given that the vacuum connection is connected to a vacuum source, this means that there is a vacuum in both chambers.

As soon as a braking operation is initiated, the piston rod moves forwards in the direction of the arrow. After a short stroke the connection between the working chamber and the vacuum chamber is blocked. As the piston rod continues to move, the inlet valve in the valve unit is opened and atmospheric air flows into the working chamber. The pressure in the working chamber is then greater than in the vacuum chamber. The atmospheric pressure acts via the diaphragm on the diaphragm disk with which it is in contact. Because the diaphragm disk is attached to the valve unit, the latter moves when the disk moves, thereby assisting the foot pressure transmitted by the connecting rod. Maximum boost is dependent on the effective diaphragm or piston area, on the atmospheric pressure and on the effective vacuum.

When the braking operation has ended, the inlet valve is closed and the vacuum and working chambers are connected via the valve unit. In this way, the pressure (vacuum) is identical in both chambers.

Vacuum non-return valve
In all brake systems which have a vacuum brake booster a non-return valve is incorporated in the vacuum line between vacuum source and brake booster. While there is a vacuum present, the non-return valve remains open. It closes when the vacuum source ceases to produce a vacuum (engine is switched off) so that the vacuum inside the brake booster is retained. Thus, brake boost is effective even when the engine is switched off for several brake actuations.

Electromechanical brake booster
The electromechanical brake booster referred to as iBooster by Bosch satisfies, thanks to its electronic control, new demands imposed on brake systems. These demands include for example lower or no availability of vacuum in the vehicle, reduced CO_2 emissions, and redundancy for highly automated driving. The iBooster can be used with all drive concepts, including hybrid and electric vehicles. Like the vacuum brake booster, the iBooster supports the driver with an assisting force (electric motor via transmission).

Operating principle
The iBooster (Figure 3) detects the driver's brake request via an integrated differential-travel sensor and sends this information to the ECU. The ECU calculates the activation of the electric motor, which converts its torque via a gear unit into the requested assisting force. Here the electric motor is activated in such a way that the differential travel between the input rod connected to the brake pedal and the transmission element connected to the electric motor is compensated to zero. The sum total of the force supplied by the booster and the driver is converted into hydraulic pressure in a standard brake master cylinder. The iBooster's resulting pedal characteristic is dependent on the design of the components (e.g., on the maximum motor force). Certain parameters can, in contrast to the vacuum brake booster, be additionally influenced by the software logic.

Special features
The iBooster enables the pedal characteristic to be adapted whereby the assisting force is adapted by means of a change to the target-value calculation of the control facility. It is thus possible to adapt the so-called jump-in and the boost factor (Figure 4) within certain limits to the requirements specified by the vehicle manufacturer. Jump-in is the point at which the driver force proportionally influences the brake-booster braking force. Below the

jump-in the braking force comes solely from the booster itself. The driver must first overcome spring forces before his/her initiated force influences the braking force.

In vehicles with an electric or hybrid drive the iBooster, in combination with a special type of driving-dynamics control (Electronic Stability Control ESC), enables brake energy to be recuperated to a vehicle deceleration of 0.3 g without affecting the braking feel. Here during braking the decelerations caused by wheel brake and electric machine are variably matched to each other without additional components. In hybrid vehicles this regenerative braking reduces fuel consumption and CO_2 emissions – especially in the event of frequent braking and accelerating in urban traffic.

The iBooster can build up brake pressure automatically (without actuating the brake pedal) with the aid of the engine/transmission unit. Compared with typical driving-dynamics control systems, the required brake pressure is built up faster and adjusted with greater accuracy. This is beneficial for example for automatic emergency-braking systems and ACC functions.

In combination with driving-dynamics control the iBooster offers the brake-system redundancy required for automated driving. Both systems are, independently of one another, able to generate brake pressure and decelerate the vehicle.

Brake master cylinder

The brake master cylinder converts the foot force applied by the driver and boosted by the brake booster into hydraulic brake pressure.

Brake master cylinder with central valve

Design

In order to comply with statutory safety requirements, service-brake systems are

Figure 4: Adaptation of pedal characteristic

1 Jump-in,
2 Pedal characteristic resulting from design,
3 Adaptation of pedal characteristic by software (adaptation of jump-in and adaptation of iBooster assisting force).

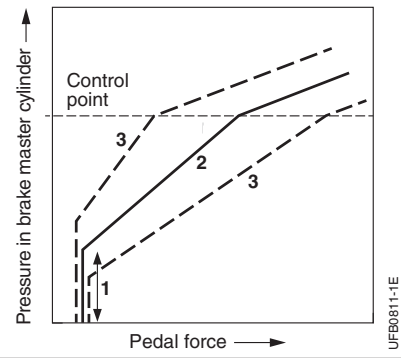
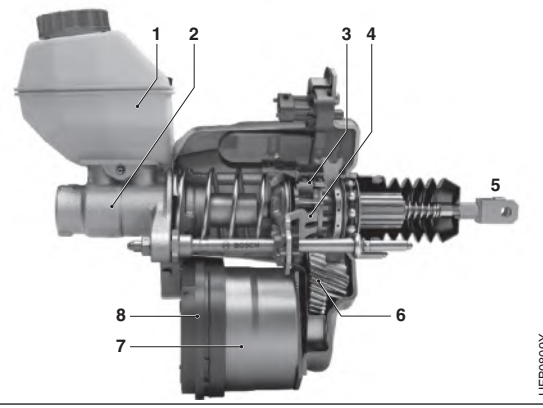


Figure 3: Electromechanical brake booster (iBooster)

1 Brake-fluid reservoir,
2 Brake master cylinder,
3 Differential-travel sensor,
4 Transmission element,
5 Input rod,
6 Transmission,
7 Electric motor,
8 Control unit.



equipped with two separate service-brake circuits. If a leak occurs (circuit failure), the other circuit remains intact (secondary braking effect). This can be achieved by a tandem brake master cylinder (Figure 5). The compression spring of the intermediate-piston circuit in the rest state holds the intermediate piston and the push-rod piston against the rear stop. The compensating port and the central valve are opened. Both hydraulic service-brake circuits are depressurized (drive position).

Operating principle
The force applied at the brake pedal and boosted by the brake booster acts directly on the push-rod piston and pushes it to the left. After short piston travel the compensating port is sealed and a pressure build-up can take place in the push-rod circuit. This also causes the intermediate piston to be pushed to the left.

Brake master cylinder with captive piston spring
Design
The "captive" piston spring – a compression spring – in the rest state holds the push-rod piston and the intermediate piston always at the same distance (Figure 6). This prevents the piston spring in the rest state from pushing the inter-

Figure 5: Tandem brake master cylinder with central valve in the intermediate-piston circuit

1 Cylinder housing, 2 Pressure chamber, intermediate-piston circuit,
3 Pressure chamber, push-rod circuit, 4 Pressure port, intermediate-piston circuit,
5 Pressure port, push-rod circuit, 6 Connection for brake-fluid reservoir, 7 Push rod,
8 Intermediate piston,
9 Central valve,
10 Stop for central valve,
11 Primary cup seal, intermediate piston,
12 Primary cup seal, push-rod piston,
13 Separating cup seal,
14 Compensating port,
15 Compression spring, intermediate-piston circuit,
16 Compression spring, push-rod circuit.

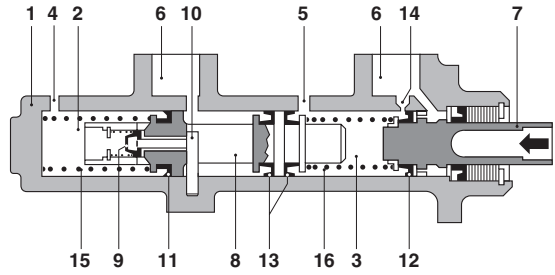
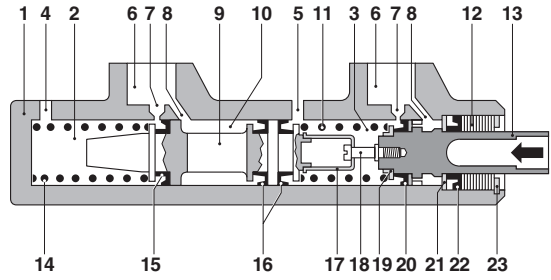


Figure 6: Tandem brake master cylinder with captive piston spring

1 Cylinder housing, 2 Pressure chamber, intermediate-piston circuit,
3 Pressure chamber, push-rod circuit, 4 Pressure port, intermediate-piston circuit,
5 Pressure port, push-rod circuit, 6 Connection for brake-fluid reservoir, 7 Compensating port,
8 Replenishing port, 9 Intermediate piston, 10 Space, 11 Captive piston spring,
12 Plastic bush, 13 Push-rod piston,
14 Compression spring, intermediate-piston circuit,
15 Primary cup seal, intermediate piston,
16 Separating cup seal,
17 Stop sleeve,
18 Stop screw,
19 Support ring,
20 Primary cup seal, push-rod piston,
21 Stop disk,
22 Secondary cup seal,
23 Snap ring.



mediate piston and the latter from overrunning the compensating port with the primary cup seal. In this situation pressure compensation via the compensating port would no longer be possible in the secondary circuit, and in the event of a residual pressure the brake shoes would not lift off the brake drums when the brake releases.

Operating principle
When the brake is actuated, the push-rod piston and the intermediate piston move in the direction of the arrow to the left, overrun the compensating ports and force brake fluid through the pressure ports into the brake circuits. As pressure increases the intermediate piston is no longer moved by the captive piston spring, but instead by the pressure of the brake fluid.

Expansion reservoir
The expansion reservoir, also called the brake-fluid reservoir, is mounted directly on the brake master cylinder and connected to it via two ports. It is both the reservoir for the brake fluid and the expansion reservoir. It compensates volume fluctuations in the brake circuits which occur after the brake is released, in response to wear of the brake linings and to temperature differences in the brake system, and during intervention by the antilock braking system (ABS) or the driving-dynamics control (Electronic Stability Control).

Transmission device
The hydraulic pressure is transmitted by the brake fluid via brake pipes as per DIN 74234 [5] and brake hoses as per SAE J 1401 [6] to the wheel-brake cylinders. Brake fluids must comply with the requirements laid down in SAE J 1703 [7] or FMVSS 116 [8] (see Brake fluids).

Wheel brakes
Floating-caliper disk brakes are usually used on the front wheels, but fixed-caliper disk brakes may also be used. Both floating-caliper disk brakes with integrated locking mechanism and Simplex drum brakes are used on the rear wheels (see Wheel brakes). Combinations of disk brakes and drum brakes (drum-in-head systems) can also be used on the rear wheels. In this case, the drum brake accommodated in the brake-disk chamber is used exclusively for the parking-brake system.

The parking-brake control device can be mechanically designed as a handbrake lever or a footbrake pedal with locking mechanism. Force is generally transmitted via cables or linkage to the wheel brakes on the rear axle. In the case of electromechanical parking brakes, the brake is actuated by means of electric motors and gearings (see Electromechanical parking-brake system).

Hydraulic modulator
Arranged between the brake master cylinder and the wheel brakes is the hydraulic modulator of the antilock braking system or the driving-dynamics control system and, depending on the scope of functions, a braking-force regulator or a braking-force limiter. These components, by limiting and adapting the brake pressure mostly on the rear axle, ensure a sensible distribution of braking force between the front and rear axles. This function can, especially in vehicles with markedly different load states, be executed on a load-sensitive basis (automatic load-sensitive braking-force metering).

The hydraulic modulator modifies the brake pressure during the braking process in such a way that the wheels are prevented from locking. Depending on the control variation, this job is performed by several solenoid valves and an electrically driven supply pump. In passenger-car brake systems, the front axle is individually controlled, i.e., each wheel is braked according to the respective grip. The rear wheels are controlled according to the select-low principle so that both rear wheels are braked together according to the wheel which has the lower grip (see also Antilock braking system and Driving-dynamics control).

Electromechanical parking brake

System overview

Conventional parking-brake systems are muscular-energy brake systems and are operated by purely mechanical means via lockable hand levers or foot pedals or via a crankgear. In electromechanical parking-brake systems, also referred to simply as electromechanical parking brakes or automatic parking brakes, the control (operating) force is generated by an electric drive.

Operation and control are effected electrically via a switch or via logic control commands by other ECUs, which enable automatic closing or opening of the parking brake. The electromechanical parking brake can only be operated when the vehicle is stationary or at low speeds (usually 3 to 15 km/h). This must also be possible when the ignition and starting switch is turned off. If electric parking-brake systems are operated at higher speeds, an emergency-braking operation is first executed by the driving-dynamics control system. The parking brake is closed when the stationary vehicle status is attained within this braking operation.

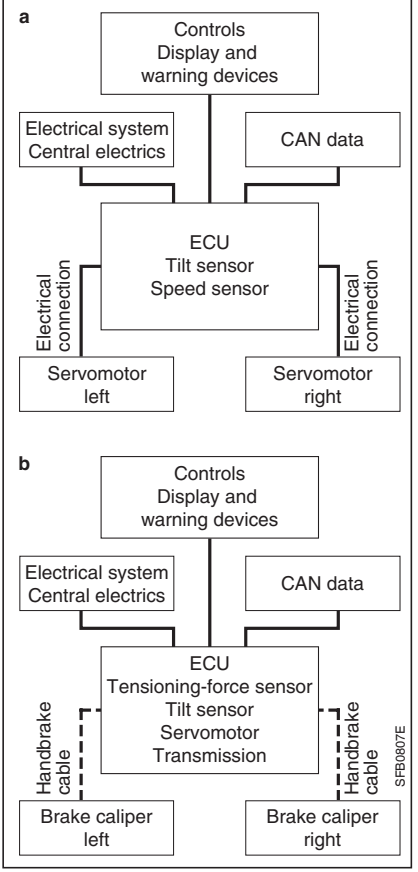
The application force in the parking mechanism (see Parking brake) depends on the slope of the gradient on which the vehicle was parked. For this purpose, depending on the system, a tilt sensor is installed in the ECU of the electromechanical parking brake or corresponding sensor signals from other ECUs are used (e.g., airbag or chassis control). The retensioning of the brake necessitated by the cooling of the mechanical brake components is performed preventatively or according to a calculated temperature model and after vehicle movement has been detected.

A safety concept must ensure that unintentional activation in both the release and the closing directions due to electrical faults is ruled out. Furthermore, intentional activation of the electromechanical parking brake (emergency braking, only necessary if the control device of the service-brake system is broken) must not give rise to critical driving situations. If the operating unit of the electromechanical parking brake is deliberately actuated on a permanent basis, the driving-dynamics control system takes over the task of braking the vehicle at a speed in excess of 10 km/h. This ensures an optimally safe braking operation even in critical road situations. The electromechanical parking brake is locked only after the vehicle speed has dropped below a specific threshold. The systems communicate with each other through an adequate data link (usually CAN or FlexRay).

Electric parking-brake systems may feature additional functions such as automatic release when starting.

Electric parking-brake systems are energy-assisted systems and are equipped

Figure 7: Electric parking-brake system
a) System with servomotor on brake caliper,
b) System with cables.



with an emergency release device. It must also be possible for the system to be activated even when the ignition and starting switch is turned off, and the system may only be released when the ignition and starting switch is turned on and the brake pedal is simultaneously being pressed (or in the event of automatic release when the brake pedal is pressed).

The driver is always alerted to a locked parking brake system by a red warning lamp. Self-diagnosis detects malfunctions and faults, and indicates them via a warning lamp. A text message can also appear in a driver information display. The fault memory can be read out with a diagnostic tester and cleared after the fault has been corrected.

Diagnosis testers and relevant software may be required for servicing work, e.g., when replacing the brake pads.

Electromechanical parking brake with servomotor on the brake caliper

The electromechanical parking brake with servomotor comprises the following components (Figure 7a): operating unit, ECU, display and warning devices, tilt sensor (can be installed in the driving-dynamics control system), floating caliper with electric motor and multiple-stage gearing. A system distribution as described in VDA Recommendation 305-100 [9] is becoming increasingly common. This entails integrating the functionality of the parking brake in the driving-dynamics control system, thereby affording freedom in the choice of different manufacturers of these systems and the parking brake.

In the case of a brake caliper with an electric servomotor, force is transmitted for the parking-brake effect via a multiple-stage gearing and a threaded spindle. It is activated by way of an electrical switch (operating unit), which forwards the control commands to the ECU redundantly and in accordance with the safety concept. The ECU, taking into account further boundary conditions (e.g., road gradient), activates the electric servomotors via separate driver stages and electrical connecting leads.

A very high gear ratio means that very high application forces can be generated. These are in the region of 15 to 25 kN. As dictated by the concept, the electromechanical and hydraulic forces can be superimposed (superposition at the brake piston).

Electromechanical parking brake with cables

In the case of an electromechanical parking brake with cables, the following components are combined in a centrally arranged assembly – above the rear axle, in the passenger compartment or in the fender (Figure 7b): electric drive motor with gearing, required sensors (depending on the scope of functions, e.g., force, tilt, temperature and position sensors), ECU, and cable mechanism (if necessary with emergency release device).

This system too is activated by way of an electrical switch, which forwards the control commands to an ECU. The ECU activates the electric servomotor(s) via a driver stage. The application force can vary, depending on the road gradient. The system is automatically retensioned when the vehicle is stopped either after a cooling phase corresponding to a temperature model or after vehicle movement has been detected.

Electrohydraulic brake

Function

The electrohydraulic brake (EHB, "Sensotronic Brake Control", SBC) is an electronic brake-control system with hydraulic actuator engineering. Like a conventional hydraulic brake, its function is to reduce vehicle speed, bring the vehicle to a halt, or keep the vehicle stationary. As an active braking system, it takes control of brake operation, braking-force boosting and braking-force control. Hydraulic standard wheel brakes are used as brakes.

Actuation unit

Mechanical operation of the brake pedal is detected by the actuation unit by means of electronic sensors with redundant backup (Figure 8). The pedal-travel sensor consists of two separate angle-position sensors. Together with the brake-pressure sensor for the pressure applied by the driver, this produces a threefold system for detecting driver input. The system can continue to function normally, even if one of the sensors fails.

The pedal-travel simulator produces an appropriate force/travel curve and calculates the amount of brake-pedal damping. The driver experiences the same "brake feel" with electrohydraulic brakes as with a very well designed conventional braking system.

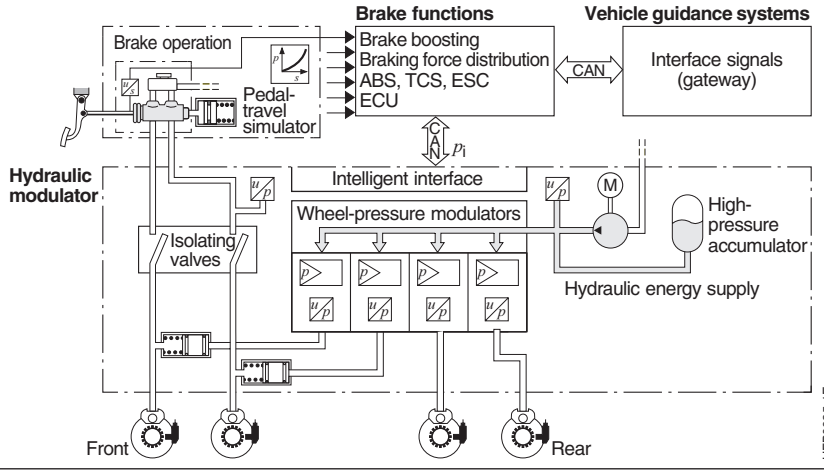
A conventional brake booster is not required here. Only the driver's brake request is determined in the actuation unit during normal operation; the brake pressure is generated in the hydraulic modulator. The brake master cylinder performs its function in the event of a system failure. The expansion reservoir supplies the hydraulic modulator with brake fluid.

Electronic control

The brake request is determined in the remote-mounted ECU from the sensor signals of the actuation unit. Braking characteristics can be adapted to the driving conditions (e.g., sharper response with more dynamic driving styles). A "duller" pedal characteristic can be used to alert the driver to a reduction in braking effect, when the brakes reach the limits of their effectiveness before overheating induces brake fading.

Further, the functions for the antilock braking system, the traction control system, and the driving-dynamics control are integrated in the ECU. In addition, comfort and convenience functions such as for example hill hold control, automatic brake-system prefilling when the foot is quickly removed from the accelerator pedal, the chauf-

Figure 8: Interaction of electrohydraulic-brake function modules



feur brake (soft stop, jolt-free stopping by means of automated brake-pressure reduction shortly before standstill), and the brake-disk wiper are provided.

Thanks to complete electronic pressure control, the electrohydraulic brake can be easily networked with vehicle guidance systems (e.g., Adaptive Cruise Control, ACC).

Hydraulic modulator

Operating principle in normal operation

Figure 8 shows the electrohydraulic-brake components as a block diagram. An electric motor drives a hydraulic pump. This charges a high-pressure accumulator to a pressure of between approximately 90 and 130 bar, monitored by an accumulator pressure sensor. The four separate wheel-pressure modulators are supplied by the accumulator and set the required pressure at the wheel-brake cylinders separately for each wheel. The pressure modulators themselves each consist of two valves with proportional-control characteristics and a pressure sensor. Brake-pressure modulation and active braking are silent and generate no brake-pedal feedback.

In normal mode, the isolating valves isolate the brakes from the actuation unit. The system is in “brake-by-wire” mode. It electronically detects the driver’s braking request and transmits it “by wire” to the wheel pressure modulators. The interaction of electric motor, valves, and pressure sensors is regulated by the ECU. This has two microcontrollers which monitor each another. The essential feature of these electronics is their extensive self-diagnosis, which monitors the plausibility of every system state at all times. It means that any faults can be displayed to the driver before a critical condition arises. If components fail, the system automatically provides the optimal remaining partial function to the driver.

An intelligent interface with the CAN bus provides the link between the remote-mounted ECU and the add-on of the hydraulic modulator.

Braking in the event of system failure

The electrohydraulic brake is designed so that in the event of serious faults (e.g., power-supply failure), it switches to a state in which the driver can brake the vehicle without using the active brake-booster function. When de-energized, the isolating valves establish a direct connection to the actuation unit and allow a direct hydraulic connection from the actuation unit to the wheel-brake cylinders (hydraulic fallback level).

Integrated power brake

Design

In response to the increasing electrification of the drivetrain and as a result the lack of vacuum supply for the brake booster, new types of braking systems have been developed in recent years. The integrated power brake (IPB, Figures 9 and 10) is vacuum-independent and combines the following components of a conventional braking system in a single unit [11]:

– connection to the brake pedal,
– brake master cylinder,
– brake-light switch,
– brake booster,
– ESC hydraulic unit (Electronic Stability Control, electronic stability program).

Further classic brake-system components such as vacuum source or vacuum pump, long hydraulic lines from the brake master cylinder to the hydraulic unit of the brake-control system, and electrical connections are rendered superfluous. The installation requirement in the vehicle is reduced as a result.

Scope of functions

The integrated power brake (IPB) contains the following functions accordingly:

– brake boosting,
– wheel-individual brake-pressure modulation for the antilock braking system (ABS), the traction control system (TCS), and driving-dynamics control (ESC, Electronic Stability Control, electronic stability program),
– external deceleration requests, e.g., from Adaptive Cruise Control (ACC),
– highly dynamic brake-pressure build-up,
– monitoring of the hydraulic brake circuits (air and leakage monitoring),
– hydraulic fallback level for fault scenarios in accordance with functional safety (ISO 26262 [12]).

Thanks to its independence from the vacuum supply, its high energy efficiency owing to the direct use of electrical energy for brake-pressure build-up (without a hydraulic high-pressure accumulator), and its capacity to recover kinetic energy as electrical energy, the integrated power brake (IPB) is suitable for use in both vehicles with electric drives (hybrid and electric vehicles) and vehicles with internal-combustion engines. It can be used with both diagonal (X-split) and front/rear brake-circuit configurations (II-split).

Figure 9: Integrated power brake (IPB)
1 Brake pedal,
2 Adapter plate on the firewall,
3 Integrated power brake (IPB),
4 Battery, 5 Wheel brakes.
A Hydraulic lines,
B Electrical power supply,
C Communication line.

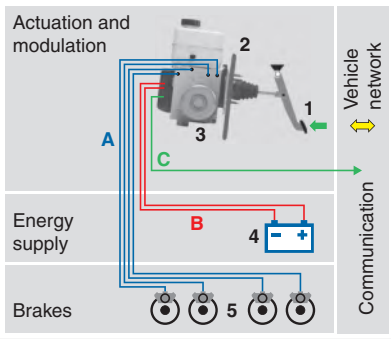
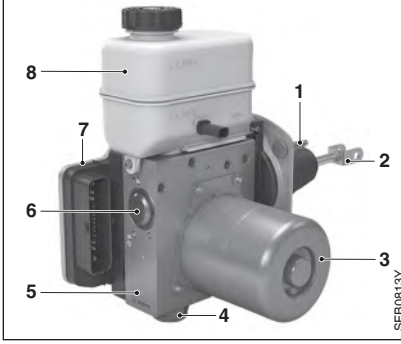


Figure 10: Components of integrated power brake (IPB)
1 Adapter for mounting on the firewall,
2 Connection to brake pedal,
3 Hydraulic module, 4 Pedal-feel simulator,
5 Valve housing, 6 Brake master cylinder,
7 Electronic control unit (ECU),
8 Brake-fluid reservoir.



The integrated power brake (IPB) is able to build up the necessary brake pressure for driver-assistance functions on the one hand very quietly and comfortably and on the other hand with high dynamic response.

The smaller number of brake-system components simplifies the production of vehicles for the vehicle manufacturer and also reduces the complexity of the value-added chain.

Functional description

Determination of the driver's braking request

The integrated power brake (IPB) is installed like a conventional brake booster on the firewall and connected with the brake pedal. The primary piston of the brake master cylinder is displaced when the brake pedal is pressed (Figure 12). The driver's braking request is determined by an integrated pedal-travel sensor.

By-wire mode

The “by-wire mode” is activated as soon as the unit detects that the driver wishes to brake (Figure 11a). The brake master cylinder is decoupled by valves from the wheel-brake cylinders and connected with a pedal-feel simulator. Because the brake pedal is decoupled from the wheel brakes the pedal feel is independent of the brake pressures at the wheels, which is beneficial to recuperation and to the configuration of the mechanical fallback level.

At the same time a plunger-cylinder system (hydraulic module) is connected via valves with the wheel brakes. The plunger of the hydraulic module is activated by an electric motor (Figure 12) precisely and in accordance with the driver's braking request. It forces brake fluid from the hydraulic module's working chamber into the wheel brakes and as a result the brake pressure requested by the driver is generated at the wheels.

Figure 12: Brake master cylinder of integrated power brake (IPB)
1 Input rod, 2 Pedal-travel sensor,
3 Primary piston, 4 Secondary piston,
5 Seals,
6 Hydraulic module with electric motor and plunger.

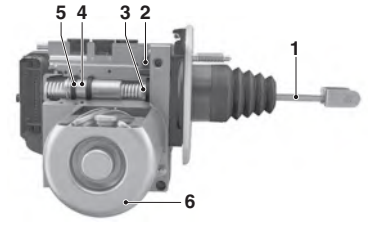
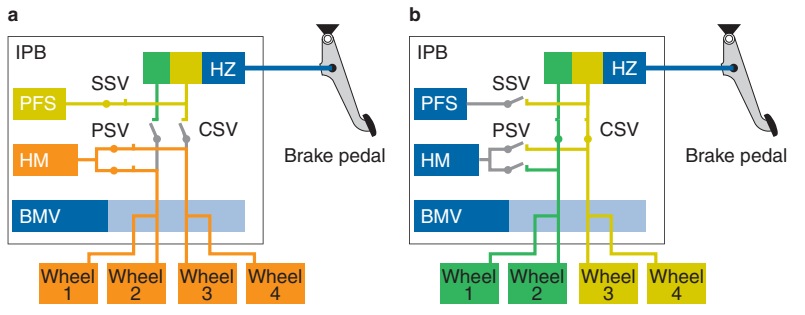


Figure 11: Active and passive operation of integrated power brake (IPB)
(Schematic diagram, separation valves are shown as “switches”).
a) Active “by-wire mode”,
b) Fallback level.
HZ Brake master cylinder, PFS Pedal-feel simulator,
HM Hydraulic module, BMV Brake-pressure modulation valves,
CSV Brake-circuit separation valve,
PSV Hydraulic-module plunger separation valve,
SSV Simulator separation valve.



Brake boosting
Brake boosting is simultaneously achieved by the pressure regulation of the hydraulic module. Because of the decoupling of the brake pedal and wheel brake pressure the brake pressure and the generator torque are masked in hybrid and electric vehicles. In other words, the braking effect of the wheel brakes can be reduced by the braking effect of the electric machine achieved by recuperation. The masking occurs continuously during recuperation and without any noticeable effects on the brake pedal.

In the case of brake-pressure requests not initiated by the driver, such as for example coming from an ACC system, pressure is generated at the wheels in the same way by means of the hydraulic module.

Wheel-individual brake interventions
For wheel-individual pressure generation or pressure modulation – similarly to an ESC system (see ESC hydraulic unit) – inlet and outlet valves (brake-pressure modulation valves, Figure 11) are integrated in the integrated power brake (IPB). A build-up of pressure at a wheel is effected via the hydraulic module when the inlet valve assigned to this wheel is opened. A reduction of pressure is effected by the opening of the outlet valve assigned to this wheel whereby the brake fluid flows directly back into the brake-fluid reservoir. The wheel-individual functions such as antilock braking system (ABS), traction control system (TCS), and lateral-dynamic stabilization by driving-dynamics control (electronic stability program).

Because the brake-fluid volume in the hydraulic module is limited by the design, the hydraulic module is recharged with brake fluid within a matter of a few milliseconds during a long ABS control operation. The brake pressure in the individual wheels is maintained during recharging and consequently there is no impact on the vehicle's braking effect.

Functional safety
In the case of safety-related vehicle components such as the braking system, safe operation and system behavior are of central importance. The electromechanical braking system (IPB) was therefore developed in accordance with the current functional-safety standards (ISO 26262).

Redundancy
Sensing of the driver's braking requests is redundantly designed in the interests of reliability. For the brake-master-cylinder piston travel the mechanism is simple and the sensor system and electronics are present in redundant form. The piston travel is also checked for plausibility against the measure brake-master-cylinder pressure.

Fallback level
To ensure that the driver is able to slow down the vehicle reliably in the event of a fault, the integrated power brake (IPB) has a direct, mechanical-hydraulic feedthrough from the brake pedal to the wheels as a fallback level (Figure 11b).

The process of decoupling the brake master cylinder from the wheel brakes (in by-wire mode) – in contrast to conventional braking systems – reduces the diameter of the brake master cylinder. In the mechanical fallback level a higher brake pressure and with it a greater deceleration of the vehicle are thus made possible with the same pedal force.

Monitoring for hydraulic changes
But, as a result of the smaller brake-master-cylinder diameter, the brake-fluid volume in the brake master cylinder is also reduced compared with a conventional braking system. The state of the braking system is therefore continuously monitored for hydraulic changes (e.g., leaks, ingress of air into the brake fluid due to leakage or faulty filling). The driver is warned before a critical system state can ensue in the event of a fault.

Reliability of the electrical power supply
The availability of brake boosting is directly allied with the supply of the integrated power brake (IPB) with electrical power. It is therefore essential to ensure that the vehicle electrical system is sufficiently reliable.

Redundancy for automated driving

Second pressure-build-up unit

For automated-driving vehicles where the driver is no longer responsible for control of the vehicle, the integrated power brake (IPB) can be enhanced by a second, independent pressure-build-up unit (RBU, redundant brake unit) to generate brake pressure. If the integrated power brake (IPB) fails, this independent unit ensures that the vehicle can be slowed down and brought to a stop.

Integrated redundancy

Integrated redundancy in the IPB may also be considered as an alternative to a second, independent pressure-build-up unit. Full redundancy is however not possible for all the system components. Seals or transmission elements for converting the rotary motion of the engine into a translatory motion of the piston cannot be redundantly designed automatically.

Market development

An integrated power brake (IPB) with integrated redundancy would result in a more complex design with increased dimensions. Automated driving is therefore only implemented with two independent pressure-build-up units, the integrated power brake (IPB) and the redundant brake unit (RBU).

References

[1] 71/320/EEC: Council Directive of 26 July 1971 on the approximation of the laws of the Member States relating to the braking devices of certain categories of motor vehicles and their trailers.
[2] ECE R13: Regulation No. 13 of the United Nations Economic Commission for Europe (UN/ECE) – Uniform provisions concerning the approval of vehicles of categories M, N and O with regard to braking.
[3] ECE R13-H: Regulation No. 13-H of the United Nations Economic Commission for Europe (UN/ECE) – Uniform provisions concerning the approval of passenger cars with regard to braking.
[4] §41 StVZO (road traffic licensing regulations, Germany) – Brakes and wheel chocks.
[5] DIN 74234: Hydraulic braking systems; brake pipes, flares.
[6] SAE J 1401: Road Vehicle Hydraulic Brake Hose Assemblies for Use with Non-petroleum-Base Hydraulic Fluids.
[7] SAE J 1703: Motor Vehicle Brake Fluid.
[8] FMVSS 116: Federal Motor Vehicle Standard No. 116: Motor Vehicle Brake Fluids.
[9] VDA Recommendation 305-100: Recommendation for integration of electric parking brakes control into the ESC system (Electronic Stability Control) with regard to the ESC (ESC assembly) and the brake caliper (brake assembly).
[10] B. Breuer, K.H. Bill (Editors): Bremsenhandbuch. 5th Edition, Verlag Springer Vieweg, 2017.
[11] U. Bauer, T. Maucher, M. Brand: Integrated Power Brake – modular set extension for highly automated driving. 8th International Munich Chassis Symposium 2017 – chassis.tech plus. Verlag Springer Vieweg, 2017.
https://rd.springer.com/book/10.1007/978-3-658-18459-9?page=2#toc
[12] ISO 26262: Road vehicles – Functional safety.

Brake systems for commercial vehicles

System overview

Brake systems for commercial vehicles and trailers must satisfy the requirements of various regulations such as, for example, RREG 71/320 EEC and ECE R13 [2]. Essential functions, effects and test methods are set out in these regulations. The entire system is subdivided into the service-brake, parking-brake, secondary-brake systems, and continuous-action brake systems.

Service-brake system

Service-brake system, tractor vehicle

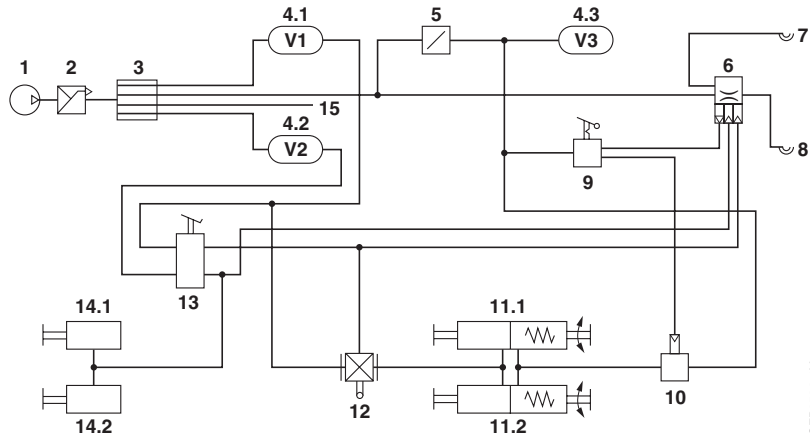
The service-brake system, designed as an energy-assisted brake system in commercial vehicles (Figures 1 and 2), can operate with compressed air or also with a combination of compressed air and hydraulics.

In the event of a fault, e.g. brake-circuit failure, it must still be possible with the operational part of the system to achieve at least the secondary braking effect – with the identical control force at the normal control device. It must be possible to meter the effect, and the trailer must not be affected by this malfunction, i.e. the trailer control (trailer control valve) must have a dual-circuit design. The secondary braking effect must achieve at least 50 % of the braking effect of the service-brake system. It is therefore customary to split the system into two brake circuits already separate on the supply side, even though this configuration is only legally required in motor buses.

The energy supply to the trailer must be guaranteed even during the braking operation. The dual-line system became mandatory when RREG 71/320 came into force, but had already been available prior to this and was known by the name “Nato” brake.

Figure 1: Structure of a compressed-air brake system with trailer control

1 Air compressor, driven by engine, 2 Pressure regulator, 3 Four-circuit protection valve,
4.1 Air reservoir V1 for circuit 1, 4.2 Air reservoir V2 for circuit 2,
4.3 Air reservoir V3 for circuit 3 (trailer, pneumatic suspension),
5 Overflow valve with limited return flow, 6 Trailer control valve with throttle valve,
7 “Supply” coupling head (red), 8 “Brake” coupling head (yellow),
9 Parking-brake valve with test position, 10 Relay valve,
11.1 Combination brake cylinder, rear right, 11.2 Combination brake cylinder, rear left,
12 Load-dependent braking-force regulator (ALB), 13 Service-brake valve,
14.1 Brake cylinder, front right, 14.2 Brake cylinder, front left,
15 Secondary loads/ancillaries (e.g. pneumatic suspension, door-closing system).



The trailer is continually supplied with a defined pressure via the supply line. This pressure must be between 6.5 and 8.0 bar on an intact tractor vehicle, irrespective of the tractor vehicle's operating pressure established by the manufacturer. The trailer must be exchangeable. The trailer's service-brake system is controlled by a second line, the brake line. This line is also governed by the regulations pertaining to trailer exchangeability. Thus the pressure in the brake line must be 0 bar in driving mode, and 6.0 to 7.5 bar in fully-braked mode.

Service-brake system, trailer

The trailer has an independent service-brake system, which is only partly subject to the demand for a secondary braking effect. According to the requirements in RREG 71/320, the braking effects of the service-brake systems in the tractor vehicle and in the trailer must be within closely set tolerances as a function of the control pressure in the brake line to the trailer, i.e. they must be approximately the same (design tolerance band RREG 71/320 and ECE R13).

If the supply line or the brake line is fractured, the trailer must be able to be fully or partly braked or it must initiate an automatic braking. Commercial vehicles with electronically controlled brake systems have – in addition to the pneumatic brake line – an electrical signal-transmission feature for electrically controlling the service-brake system in the trailer. This is

Figure 3: Compatibility diagram

Tractor vehicle and trailer RREG 71/320, ECE R13.

T_R Sum total of braking forces at circumference of all trailer wheels,
P_R Total static normal force of trailer,
T_M Sum total of braking forces at circumference of all tractor-vehicle wheels,
P_M Total static normal force of tractor vehicle,
P_m Pressure at "Brake" coupling head.

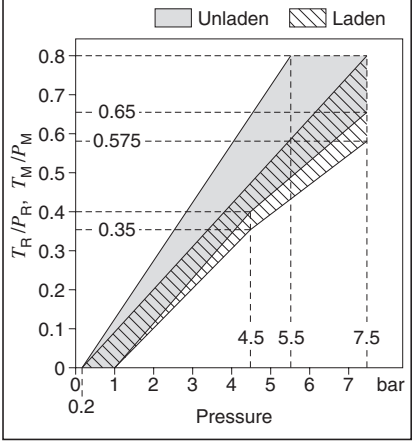
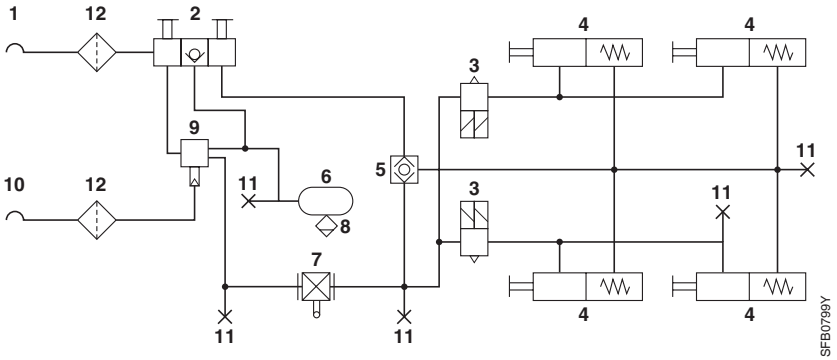


Figure 2: Compressed-air brake system of a 2-axle semitrailer with ABS (simplified representation)

1 "Supply" coupling head (red), 2 Double release valve, 3 1-channel ABS pressure-control valve, 4 Combination brake cylinder, 5 Shuttle valve, 6 Air reservoir, 7 Load-sensitive braking-force regulator, 8 Drain valve, 9 Trailer brake valve, 10 "Brake" coupling head (yellow), 11 Test connection, 12 Line filter.



performed by a standardized electrical plug-in connection in accordance with ISO 7638 [3]; this plug-in connection may have 5 or 7 pins.

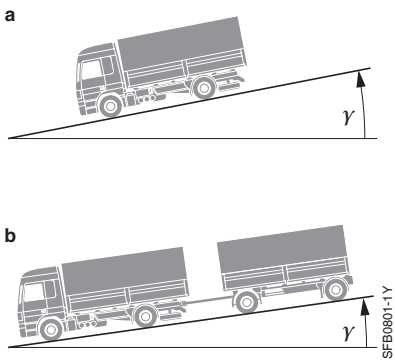
Tractor vehicles and trailers must be freely exchangeable. Compatibility conditions have therefore been defined in Annex 2 RREG 71/320 and ECE R13. Accordingly, the ratio between retardation and pressure at the "Brake" coupling head in the range depicted in Figure 3 must be in the range of 0.2 to 7.5 bar at the "Brake" coupling head. The diagram only applies to tractor vehicle and trailer. All other vehicles and vehicle combinations are covered by other diagrams.

Parking-brake system

The parking-brake system is an independent brake system which must hold the vehicle stationary after it has been brought to a stop, even when the driver is not in the vehicle. The holding-stationary effect is calculated on a downhill gradient with fully laden vehicles. The downhill gradient for Category M, N, O single vehicles (except O1) is 18%. If the vehicles are equipped to tow a trailer, the holding-stationary effect must also be achieved with an unbraked trailer. The downhill gradient is then only 12% (Figure 4).

Figure 4: Test condition for parking-brake system

a) Single vehicle, 18% downhill gradient.
b) Tractor vehicle and trailer, 12% downhill gradient; only the tractor vehicle is braked.
γ Gradient angle.



The parking-brake system in commercial vehicles and motor buses is usually designed as a spring-loaded brake system. The springs in the spring-type brake cylinders, when the wheel brakes are adjusted in accordance with the regulations, generate the same force as the pneumatic brake cylinders in the service-brake system when the nominal pressure (brake-system calculated pressure) acts on their nominal effective area. If certain malfunctions occur, e.g. brake-circuit failure or energy-source failure, the spring-loaded brakes may not brake automatically and must therefore be protected and designed accordingly.

Energy-assisted parking-brake systems (spring-loaded brake systems) must be equipped with at least one emergency release device. This may be a mechanical, pneumatic or hydraulic device. The parking-brake system must then only be designed for graduated (metered) operation if it has to be engaged to achieve the prescribed secondary braking effect.

In the trailer, the parking-brake system frequently operates as a muscular-energy brake system. If the trailer control system in the tractor vehicle is configured in such a way that the service brake in the trailer also responds when the parking brake is operated in the tractor vehicle (trailer control valve with port 4.3), the parking-brake valve must be provided with a test setting. This makes it possible to release the service brake in the trailer when the parking brake in the tractor vehicle is operated. This in turn makes it possible to check whether only the tractor vehicle braked with the parking-brake system can hold the entire vehicle combination.

Secondary-brake system

There is no independent secondary-brake system. It comes into effect if a malfunction, e.g. brake-circuit failure or energy-source failure, occurs in the service-brake system. In this event, it must still be possible for at least two wheels (not on the same side) to be braked.

The brake system in the trailer too must not be affected by this malfunction. For this reason, the brake systems and activation of the trailer have dual-circuit designs.

The supply volume must be designed such that, in the event of an energy-source failure, after eight full operations of the service brake there is still enough pressure available to achieve the secondary braking effect on the ninth full operation. In the event of a failure of a brake circuit on the supply side, it is essential to ensure that, when the energy source is intact, the pressure in the intact brake circuits does not drop permanently below the nominal pressure. This is achieved by using special protective devices, e.g. a four-circuit protection valve or an electronic unit.

Continuous-operation brake system
The wheel brakes used are not designed for continuous operation. Prolonged braking (e.g. on hill descents) can result in the brakes being thermally overloaded. This causes a reduction in braking effect ("fading") or, in extreme cases, in complete brake-system failure.

A wear-free brake system is referred to as a continuous-operation (retarder) brake system. In Germany, they are required in accordance with StVZO §41 s. 15 [4] for motor buses with a gross weight rating of more than 5.5 t and for other motor vehicles with a gross weight rating of more than 9 t. A retarder must hold a fully laden vehicle over a distance of 6 km and a downhill gradient of 7% at a speed of 30 km/h.

The service brake must be designed accordingly for trailers. Operation of the retarder in the tractor vehicle must not cause the service brake in the trailer to be operated (see also StVZO §72 [5]).

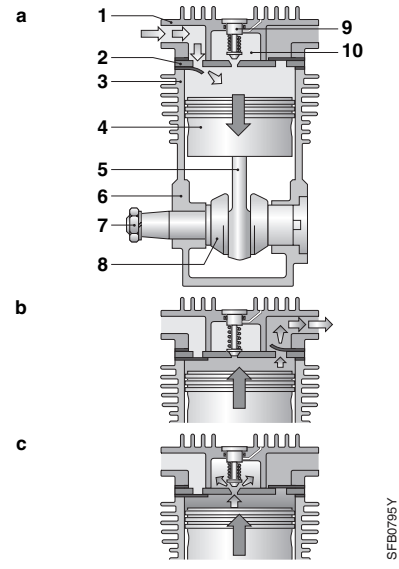
Components of commercial-vehicle brake systems

Air supply and air processing
Air supply and processing comprises the energy source, pressure control, air processing, and compressed-air distribution.

Compressor
A compressor is the source of energy. It takes in air and compresses it to compressed air, the working medium for the brake systems and the ancillaries (e.g. pneumatic suspension, door-closing system).

The compressor is a plunger pump where the crankshaft is driven directly by the vehicle engine (Figure 5). It is fitted to

Figure 5: Compressor
a) Induction,
b) Compression and discharge,
c) Compression into the additional clearance.
1 Cylinder head, 2 Intermediate plate (with inlet and outlet valves), 3 Cylinder, 4 Piston, 5 Connecting rod, 6 Crankcase, 7 Drive 8 Crankshaft, 9 ESS valve (Energy-Saving System), 10 Additional clearance.



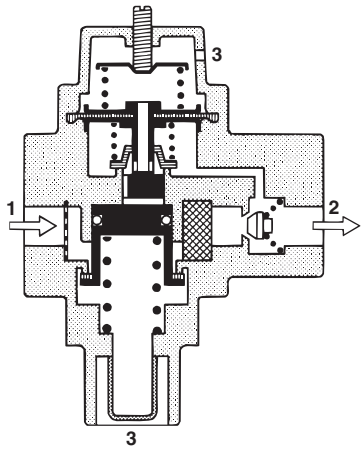
the vehicle engine by means of a flange. Its component parts are:
– The crankcase, which forms a monobloc unit with the cylinder. It contains the crankshaft with connecting rod and piston.
– The cylinder head with intake and pressure connections, as well as connections for water cooling.
– The intermediate plate with inlet and outlet valves.

To reduce losses in the idle mode (opening and flow resistances in valves and lines), an energy-saving system (ESS) is used; this activates a clearance and thus reduces compression work. This reduces fuel consumption.

During its return stroke, the piston draws in air after the inlet valve opens automatically due to vacuum. The inlet valve closes at the start of the piston's return stroke. In the fore stroke, the piston compresses the air. When a certain pressure is reached, the outlet valve opens and compressed air is supplied to the brake system.

Nowadays, compressors achieve a stroke displacement of up to 720 cm^3, a pressure level of up to 12.5 bar, and a maximum speed of 3,000 rpm. Their features include high efficiency, low oil consumption, and a long service life.

Figure 6: Pressure regulator
1 From compressor,
2 To the compressed-air reservoirs,
3 Venting.



Pressure regulator
The pressure regulator controls the compressed air supplied by the compressor in such a way that the operating pressure lies within the activation and cutoff pressure (Figure 6).

As long as the pressure in the compressed-air reservoirs lies below the cutoff pressure, connections 1 and 2 are connected and compressed air passes the pressure regulator. Once the cutoff pressure is reached, the pressure regulator switches to idle mode. Here, the venting piston is activated and connection 1 is connected to the atmosphere (venting).

Air drier
The air drier cleans the compressed air and dries it to prevent corrosion and freezing in the brake system during winter operation.

Basically, an air drier consists of a desiccant box and a housing. The housing incorporates the air passage, a bleeder valve, and a control element for granulate regeneration (Figure 7). The granulate is regenerated by activating a regeneration-air tank.

When the bleeder valve is closed, compressed air from the compressor flows through the desiccant box and from there to the supply-air reservoirs. At the same time, a regeneration-air tank is filled with dry compressed air. As compressed air flows through the desiccant box, water is removed by means of condensation and adsorption.

The granulate in the desiccant box has a limited water absorption capacity and must therefore be regenerated at regular intervals. In the reverse process, dry compressed air from the regeneration-air reservoir is reduced to atmospheric pressure via the regeneration throttle upstream of the air drier, flows back through the moist granulate from which it draws off the moisture, and flows as moist air via the open bleeder valve to the atmosphere.

The pressure regulator and air drier can be combined into one unit.

Four-circuit protection valve

The four-circuit protection valve distributes the compressed air to the various brake and ancillary circuits, isolates the circuits from one another, and ensures the water supply for the remaining circuits in the event of failure of a circuit (Figure 8).

The function of the four-circuit protection valve is provided by overflow valves specially developed for this application. In contrast to a normal overflow valve, the design of this type of overflow valve features two different effective areas on the intake-flow side. The incoming pressure from the pressure regulator acts on one effective area, while the pressure available in the pneumatic circuit acts on the other. The opening pressure of the overflow valves is thus dependent on the (residual) pressure of the assigned pneumatic circuit.

The overflow valves can be arranged differently. Often service-brake circuits 1 and 2 and ancillary circuits 3 and 4 are connected in pairs in succession. This ensures that at least one of the two service-brake circuits is filled as a matter of priority.

Figure 7: Air drier with integrated pressure regulator

1 Desiccant box, 2 Compression spring,
3 Desiccant, 4 Cup (control valve),
5 Compression spring, 6 Pin,
7 Diaphragm, 8 Compression spring,
9 Heating element, 10 Bleeder valve,
11 Drain connection, 12 Throttle,
13 Non-return valve, 14 Preliminary filter,
15 Secondary filter.

Ports:
1 From compressor,
21 To air reservoir,
22 To regeneration-air tank,
3 Vent.

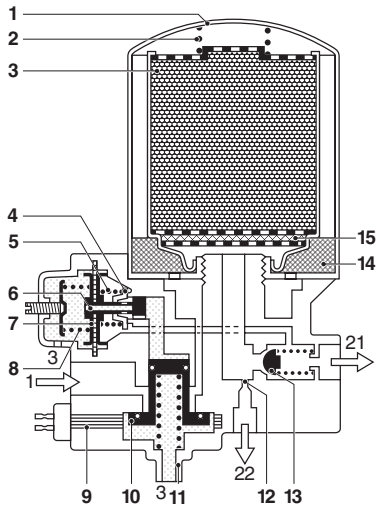
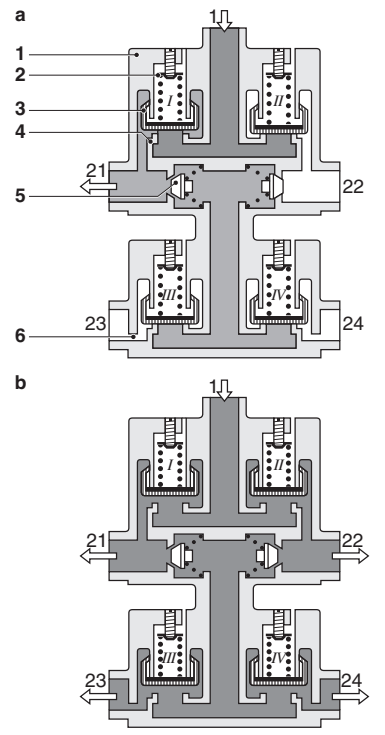


Figure 8: Four-circuit protection valve

a) Filling a compressed-air reservoir,
b) Filling all compressed-air reservoirs.
1 Housing,
2 Compression spring,
3 Diaphragm piston,
4 Valve seat,
5 Non-return valve,
6 Fixed throttle.

I – IV overflow valves

Ports:
1 Energy input,
21 – 24 Energy output to circuits 1 – 4.



The ancillary circuits for this type of valve are additionally protected by two non-return valves. These non-return valves can be omitted from four-circuit protection valves with central intake flow. These overflow valves may also be provided with variable flow restrictors. These enable an empty system to be filled with small quantities of air.

If a malfunction occurs for example in circuit 1 (circuit failure due to a leak), the pressure drops initially only in circuit 1 to 0 bar and in circuit 2 to the closing pressure. The pressure in circuits 3 and 4 is initially maintained by the effect of the non-return valves, but will also drop through consumption to the closing pressure. The intact circuits continue to be supplied under subsequent delivery by the compressor, because the residual pressure in circuits 2, 3 and 4 acts on the secondary effective area of the corresponding overflow valves. The intact circuits are filled again until the opening pressure of the defective circuit (circuit 1) acts on the primary effective area of the corresponding overflow valve, opening this valve. A further pressure increase is not possible, because from this moment the delivered compressed air is lost through the defective circuit. The opening pressure via the primary effective area is adjusted in such a way that it is equal to or above at least the nominal pressure (calculated pressure) of the brake system. This ensures both a sufficient supply of compressed air for the intact service-brake circuit and the secondary braking effect. The supply to the ancillaries such as, for example, trailer, parking-brake system and pneumatic suspension is also maintained.

Electronic air-processing unit

Nowadays, the pressure control, air processing, and compressed-air distribution are combined in one electronic unit, the air-processing unit. The electronic air-processing unit (EAC, Electronic Air Control) is a functional agglomeration of the pressure regulator, air drier and multiple-circuit protection valve into one mechatronic device. In addition, the control system of the parking brake is integrated in part. In all, integrating many functions in a mechatronic unit offers significant benefits with regard to system expenditure, functionality, and energy saving.

Energy storage

The energy required for the braking operation and for the function of the ancillaries is provided and stored in sufficient quantities in compressed-air reservoirs approved for use in road vehicles. The volume must be designed such that, without subsequent delivery, after eight full brakings the secondary braking effect prescribed for this vehicle is still achieved at least by the ninth full braking.

Despite the use of an air drier, the compressed-air reservoirs are equipped with manual or automatically acting drain devices. Compressed-air reservoirs are subject to the requirements of § 41a s. 8 [4] in conjunction with § 72 StVZO [5], and must be approved for use and permanently identified.

The supply systems for the brake systems must be fitted with warning devices. The following requirements apply:
– Red warning light,
– visible to the driver at all times
– comes on no later than on brake application or if the pressure in the supply system for the service brake has dropped to 65 % nominal pressure. 80 % nominal pressure applies to the supply system for the parking-brake system (spring-loaded brake).

Service-brake valve

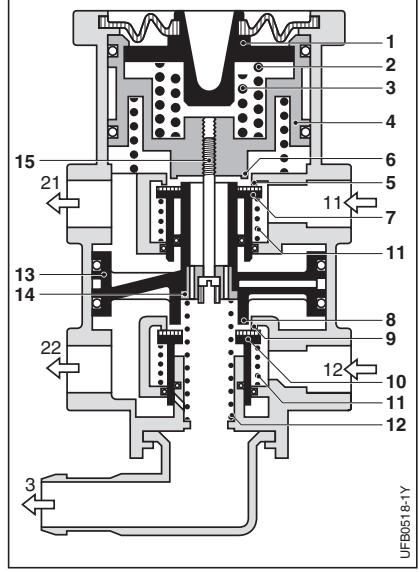
Service-brake valves (Figure 9) have a dual-circuit design and control the service-brake circuits according to the control force (force-controlled valves).

Circuit 1 is actuated by the control device, the push rod and the compression springs (travel-compensating springs). The reaction piston is forced downwards, first closing the outlet valve and then opening the inlet valve. Compressed air is admitted into brake circuit 1 and the pressure increases. The brake pressure acts in the upward direction against the reaction piston, forcing it against the compression springs as long as the partial braking range is not exceeded. The brake end position is reached, with an equilibrium of forces existing at the reaction piston.

Circuit 2 is controlled by the brake pressure in circuit 1. This acts, instead of the control device from above, on the reaction piston of circuit 2. At approximately the same time, the brake end position is reached in circuit 2 as well. In the full-braking position or in the event of a failure of circuit 1, both reaction pistons are mechanically moved to their full extent by means of the control device. The outlet valves are closed, and the inlet valves remain open. Circuits 1 and 2 are pneumatically fully and safely isolated from each other.

Special designs facilitate different controlled brake pressures for circuits 1 and 2. These are required if a dual-circuit booster cylinder is actuated by the service-brake valve, or if circuit 2 is subject to load-sensitive control. This is made possible by installing an appropriate spring assembly or a reaction piston with several effective areas.

Figure 9: Service-brake valve
1 Push rod,
2 and 3 Compression springs,
4 Reaction piston,
5 and 9 Inlet-valve seat,
6 and 8 Outlet-valve seat,
7 and 10 Valve plates,
11 Valve springs,
12 Return spring,
13 Control plunger,
14 Spring seat,
15 Connecting rod.
Ports:
3 Vent,
11 Energy input, circuit 1,
12 Energy input, circuit 2,
21 Brake pressure, circuit 1,
22 Brake pressure, circuit 2.



Parking-brake valve

Parking-brake valves (Figure 10) control the pressure in the spring-type brake cylinders as a function of the lever travel (travel-controlled valves). The lever must be permanently and securely lockable in the brake-applied position. Parking-brake valves must only operate under graduated (metered) application when the effect of the parking brake is required to achieve the secondary braking effect. Parking-brake valves must be provided with a test setting when the service-brake system is activated in the trailer on actuation of the parking brake.

Different variants of parking-brake valve exist, depending on their application: non-graduated, graduated or graduated with steep-droop characteristic curve. The latter variant provides for a highly sensitive graduable effect, because the operating range of spring-type brake cylinders, considered over the lever angle of the parking-brake valve of approximately 80°, is optimally used. The operating range of spring-type brake cylinders is between approximately 5 bar (start of braking) and approximately 2 bar (end of braking, see diagrams in Figure 11).

In pneumatic high-pressure brake systems (operating pressure greater than 10 bar), the parking-brake valve can be fitted with a pressure limiter so that standard spring-type brake cylinders can be used. The facility in parking-brake valves for attaining the capability of metering the controlled pressure is similar to the facility in service-brake valves, but operates in the opposite direction, because the spring-type brake cylinders are ventilated

Figure 10: Parking-brake valve (driving mode)

1 Actuating lever, 2 Detent element,
3 Eccentric element, 4 Return spring,
5 Outlet valve seat, 6 Inlet-valve seat,
7 Valve plate, 8 Reaction piston,
9 Reaction spring, 10 Compression spring.
Ports:
1 Energy input, circuit 3,
2 Control pressure to parking brake,
3 Vent.

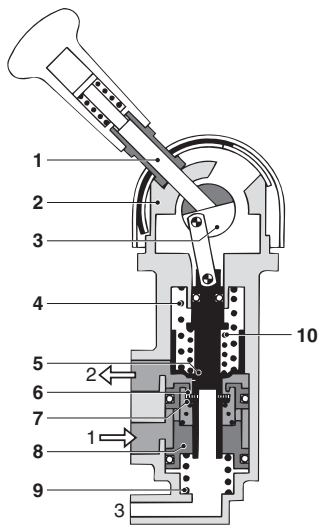
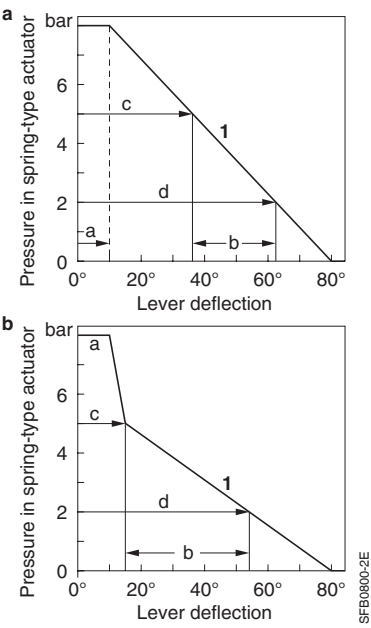


Figure 11: Working range of spring-type brake cylinder

a) Normally metered parking-brake valve,
b) Metered parking-brake valve with steep-droop characteristic curve.
1 Pressure characteristic.
a Free travel (valve lift), b Actuation range,
c Start of braking, d End of braking.



in driving mode and the brake-applied mode is achieved by bleeding.

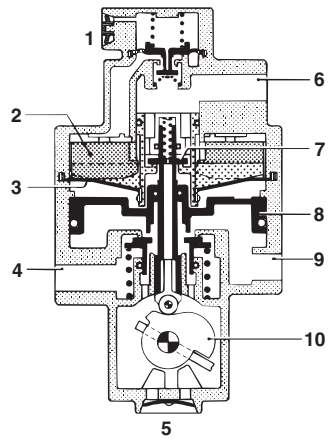
Parking-brake valves can have a dual-circuit design. The system is supplied in this case from circuit 3 and – the pneumatic auxiliary release device of the spring-type actuators – from circuit 4. An additionally required rotary-knob, shuttle or check valve can be omitted.

In the version with steep-droop characteristic curve (Figure 11), the start of braking is attained earlier and the actuation range is significantly greater. This is particularly advantageous when the parking brake is used as a secondary brake.

As an alternative to the pneumatic parking-brake valve, there are electronically controlled parking-brake systems (Electronic Parking Brake, EPB). These consist of an EPB module (which can optionally also be integrated in the air-processing unit) and an operating unit. The EPB module contains a bistable valve that can be controlled via integrated solenoid valves and further solenoid valves for executing the trailer test function.

Figure 12: Braking-force regulator with relay valve

1 Vent, 2 Rake, 3 Transfer diaphragm, 4 Energy input from air reservoir, 5 Vent, 6 Uncontrolled pressure from service-brake valve, 7 Control valve, 8 Relay piston, 9 Controlled brake pressure to brake cylinders, 10 Rotary cam.



The parking brake can be applied and released using the operating unit, whereby the last active state is retained even after the supply voltage is shut off. Since the operating unit must have a graduable design, the parking brake can also be operated in stages and consequently function as a secondary brake.

As well as manual activation by operating unit, the EPB also provides for a series of comfort and convenience functions, e.g., automatic application when the vehicle is at a standstill (Autopark) and automatic release when starting (Autorelease).

Automatic load-sensitive braking-force regulator

A facility frequently used in the service-brake system of commercial vehicles with pneumatically controlled service-brake systems is automatic load-sensitive braking-force control (ALB). Valves that handle the braking-force distribution enable adjustment of the braking forces to the low axle loads in the partially laden and unladen state and thus a correction of the braking-force distribution on the axles of an individual vehicle or a certain braking level in road trains or semitrailers.

The braking-force regulator (Figure 12) is connected between service-brake valve and brake cylinder. Depending on vehicle payload, it regulates the applied braking pressure. The device has a transfer diaphragm with variable effective area. The diaphragm is held in two radially arranged, interlocking rakes. Depending on the vertical position of the control-valve seat, there is a large reaction area (valve position at bottom) or a small reaction area (valve position at top). Consequently, the brake cylinders are supplied via an integrated relay valve with a reduced pressure which is lower than (unladen), or which is the same as (fully laden) the pressure coming from the service-brake valve. The control-valve can be moved into the load-sensitive position by means of an eccentric element that is connected via linkage to the vehicle axle or by means of a wedge (in the case of vehicles with pneumatic suspension).

The pressure limiter which is integrated into the device at the top allows a small

partial pressure (approx. 0.5 bar) to flow in to the top of the diaphragm. Thus, up to this pressure there is no reduction in brake-cylinder pressure. This results in the synchronous application of the brakes on all vehicle axles.

As an alternative to the pneumatic ALB valve, pneumatically controlled brake systems are increasingly making use of the EBD function (Electronic Brakeforce Distribution) of ABS, with which braking-force distribution is optimized as a function of the wheel slip (see Wheel-slip control systems).

The electronically controlled brake (EBS) has become the predominant system used in heavy European commercial vehicles. This system performs braking-force distribution electronically as a function of the laden state and other parameters (see Electronically controlled brake system)

Combination brake cylinder

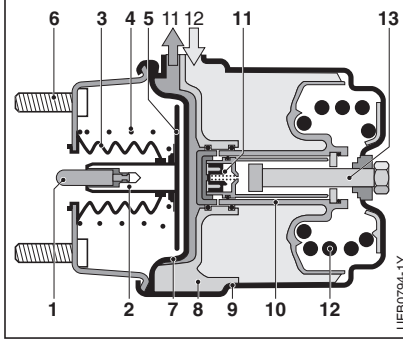
The combined cylinder in the commercial vehicle consists of a diaphragm cylinder part for the service brake and a spring-type actuator part for the parking brake (Figure 13). They are arranged one behind the other and exert force on a joint push rod. A distinction can be made between combined cylinders for S-cam brakes, wedge-actuated brakes, and disk brakes based on the type of wheel brake.

The two cylinders can be actuated independently of one another. Simultaneous actuation results in the addition of their forces. This can be prevented by installing a special relay valve in order to automatically prevent mechanical overloading of other downstream components (e.g. brake drums).

A central release screw allows for a tensioning of the spring of the spring-type brake cylinder without compressed air having to be applied (mechanical emergency release device). This is necessary to assist fitting or, in the event of failure of the compressed air, to be able to maneuver the vehicle.

Figure 13: Combination brake cylinder for disk brake (driving mode)

1 Pressure pin, 2 Piston rod,
3 Bellows with seal to disk brake,
4 Compression spring (diaphragm cylinder),
5 Piston (diaphragm cylinder),
6 Housing with fastening bolts,
7 Diaphragm, 8 Intermediate flange,
9 Cylinder housing (spring-type brake actuator),
10 Piston (spring-type brake actuator),
11 Bleeder valve (spring-type brake actuator chamber),
12 Compression spring (spring-type brake actuator),
13 Release device (spring-type brake cylinder).
Air ports: 11 Service brake, 12 Parking brake.



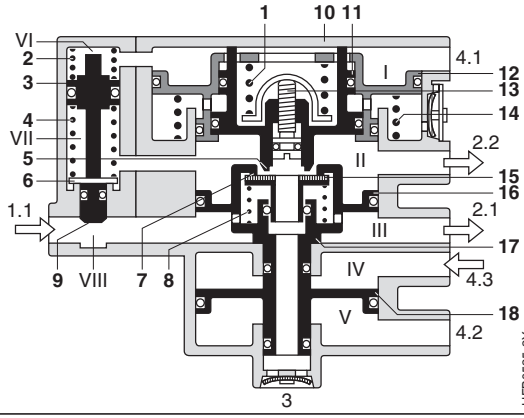
When the service brakes are operated, compressed air flows into the diaphragm cylinder and presses the plunger disk and the push rod against the lever in the disk brake. A drop in air pressure releases the brake.

When compressed air flows into the spring-type actuator part, the piston presses the springs together and the brake is released. If the chamber is vented, the spring-type brake cylinder exerts a force via the piston rod on the diaphragm part and presses the push rod into the mechanism of the disk brake via the piston disk.

Figure 14: Trailer control valve with decoupling function (driving mode)
1 and 2 Compression spring, 3 Control plunger, 4 Spring assembly, 5 Outlet-valve seat, 6 Disk, 7 Inlet-valve seat, 8 Compression spring, 9 Throttle pin, 10 Housing, 11 and 12 Control plunger, 13 Adjusting screw, 14 Compression spring, 15 Valve disk, 16 Reaction piston, 17 Collar, 18 Control plunger.

I – VIII Chambers.

Ports:
1.1 Energy input from circuit 3,
2.1 Energy output to "Supply" coupling head (red),
2.2 Energy output to "Brake" coupling head (yellow),
4.1 Control port uncontrolled pressure circuit 1,
4.2 Control port uncontrolled pressure circuit 2,
4.3 Control port parking brake,
3 Central vent.



Trailer control valve

The trailer control valve installed in the tractor controls the trailer's service brake. This multi-circuit relay valve is triggered by both service-brake circuits and by the parking brake (Figure 14). In the driving mode, supply chamber III and chamber IV of the parking-brake circuit are under the same pressure. The brake line to the trailer is connected to the atmosphere via the central venting. A pressure increase in chamber I of brake circuit 1 and in chamber V of brake circuit 2 leads to the corresponding pressure increase in chamber II for the brake line to the trailer. A pressure drop in both brake circuits also leads to the same pressure drop in the brake line. Operation of the parking-brake system leads to venting of the parking-brake circuit (chamber IV). This increases the pressure in chamber II for the brake line to the trailer. When air enters chamber IV, the brake line is vented again.

If the brake line to the trailer is pulled off, it is prescribed that the pressure in the supply line to the trailer must have fallen to a pressure of 1.5 bar in less than two seconds (RREG 71/320). To achieve this, the compressed-air supply to the supply line is throttled by means of an integrated valve.

Electronically controlled brake system

Requirements and functions

With the further development of the dual-line compressed-air brake system, an electronic (or electronically controlled) brake system (EBS) was created in the mid-1990s. On account of the modular design, it is possible to cover different vehicle types with just a few components. Vehicle-specific differences and factors can be covered to a large extent by programming the central ECU accordingly. The control arrangement is determined by the number of axles and their arrangement and the required scope of functions, and ranges from 4S/4M to 8S/6M (S wheel-speed sensor, M pressure-control module).

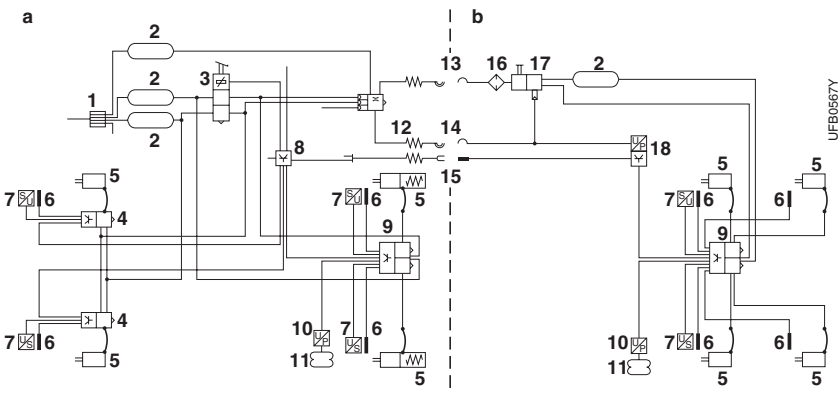
Design and operating principle

The electronic brake system (Figure 15), like a conventional compressed-air brake with antilock braking system (ABS), comprises a compressed-air supply system, but the pressure-regulator, air-drier and multiple-circuit protection-valve functions can be combined in an electronic unit (EAC, Electronic Air Control). It is thus possible to adapt certain functions – such as e.g. fill sequences or regeneration – better to the required conditions and to guarantee even greater functional reliability.

In the electronic brake system too, energy is stored in compressed-air reservoirs and from there made available to the pressure-control modules and to the service-brake valve. The service-brake valve consists of an electric pedal-travel sensor and a pneumatic section which is functionally identical to the previous design. The pedal-travel sensor consists of a redundant travel sensor (e.g. two redundantly arranged electrical potentiometers or a redundant inductive travel sensor) which is deflected via the control device and delivers the redundant output signal to the central ECU. From this the ECU cal-

Figure 15: Service-brake system of an electronically controlled braking system

a) Tractor vehicle, b) Trailer.
1 Four-circuit protection valve, 2 Air reservoir, 3 Service-brake valve with braking-level sensor,
4 Single-channel pressure-control module, 5 Brake cylinder, 6 Wheel-speed sensor,
7 Brake-lining wear sensor, 8 EBS ECU in tractor vehicle,
9 Two-channel pressure-control module, 10 Pressure sensor, 11 Air-spring bellows,
12 Trailer control valve, 13 "Supply" coupling head (red),
14 "Brake" coupling head (yellow), 15 ISO 7638 plug-in connection (7-pin), 16 Line filter,
17 Trailer brake valve with release device, 18 EBS ECU in trailer.



culates for each wheel an individual brake pressure and in turn activates the pressure-control modules on the individual axles so that the required brake pressure is applied to the brake cylinders downstream of the pressure-control modules. The applied brake pressure is regulated with the aid of integrated pressure sensors in the pressure-control modules. A pneumatic brake-pressure generation is performed in parallel in the pneumatic part of the service-brake valve which on the one hand determines the brake feel and on the other hand acts as a fallback level in the event of an electrical fault.

The brake-pressure modules are available as one-channel or two-channel designs. If the vehicle is set up to tow a trailer, a trailer control module is also provided as a substitute for the trailer control valve. This trailer control module is likewise activated in the braking operation by the central ECU and makes an adapted control pressure available at the “Brake” coupling head (yellow). This makes it possible to carry a conventionally braked trailer. If a trailer is carried with an independent electronic brake system, this is controlled by an electrical connection via the plug-in connection as per ISO 7638 (ABS connector). The trailer must nevertheless also be pneumatically coupled, because this is the only way that the trailer can be supplied with pressure and in the event of a system failure pneumatically controlled. By controlling the electronic brake system in the trailer, it is possible to provide optimum matching with regard to braking performance between the tractor vehicle and the trailer. Simultaneous and matched braking performance facilitate optimized coupling-forcing matching.

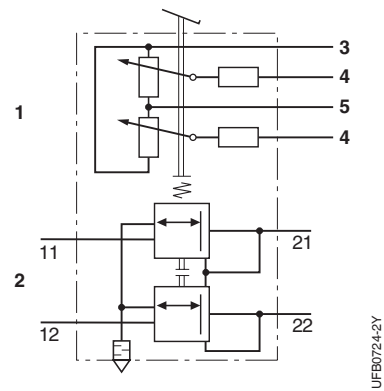
Further functions, such as antilock braking system (ABS), traction control system (TCS) and driving-dynamics control system (Electronic Stability Program), are integrated into the electronic brake system’s scope of functions. The turning behavior of the wheels is monitored by the wheel-speed sensors and the antilock braking system. Depending on the design, the information is made available to the central ECU or to the pressure-control module, where it is processed. In the event of incipient wheel locking, depending on the system arrangement and design, a control intervention is effected via the pressure-control modules or by downstream pressure-control valves in accordance with the control variants known to the ABS system (individual control, modified individual control, or select-low control). Intervention by the traction control system when the wheels spin takes the form of an engine and brake intervention. Further sensors are needed for the functions of the driving-dynamics control system. The steering-wheel angle is recorded by a steering-angle sensor. A yaw-velocity sensor, also known simply as a yaw sensor, records the rotational speed about the vehicle vertical axis. A lateral-acceleration sensor also records the lateral acceleration. When the data have been evaluated, swerving or jackknifing is detected and stabilized by the specific introduction of brake pressure into the relevant brake cylinders and intervention in other systems (see Driving-dynamics control for commercial vehicles).

If an electrical fault occurs, the vehicle can be braked by means of one or two redundant pneumatic circuits with at least the demanded secondary braking effect, and the trailer brake system controlled.

Optimum cooperation between all the systems can be achieved through data communication with other systems in the vehicle and the trailer. Optimized deceleration and acceleration processes and additional functions can be realized in this way.

Advantages of electronic brake systems in a commercial vehicle are:
– Fast and simultaneous brake-pressure buildup in all the brake cylinders,
– good metering capability, thus optimum braking comfort,
– optimum matching between tractor and trailer through control of coupling forces,
– exact braking-force distribution,
– uniform brake-lining wear,
– ABS, TCS and ESC functions are integrated (brake and engine interventions), traction control for offroad applications can be easily realized,
– driving-dynamics control via engine and brake interventions when oversteering or understeering is detected, in response to the risk of swerving (articulated road train, articulated bus), and intervention in response to the risk of overturning,
– ease of servicing thanks to extensive diagnosis functions.

Figure 16: Service-brake valve with two pneumatic control circuits
1 Braking-level sensor,
2 Service-brake valve,
3 Supply connection,
4 Electrical potentiometer connections,
5 Electrical ground connection.
Pneumatic ports:
11 Energy input, circuit 1,
12 Energy input, circuit 2,
21 Back-up control pressure, circuit 1,
22 Back-up control pressure, circuit 2.



Components of the electronic brake system

Electronic control unit (ECU)
The control point of an electronic brake system consists of a central ECU in which all the system functions run. Aside from the cab-mounted variants, frame-mounted ECUs are increasingly gaining acceptance. The benefit of the latter is that the sensors for yaw rate and lateral acceleration can be integrated and consequently there is no need for a separate sensor to be fitted.

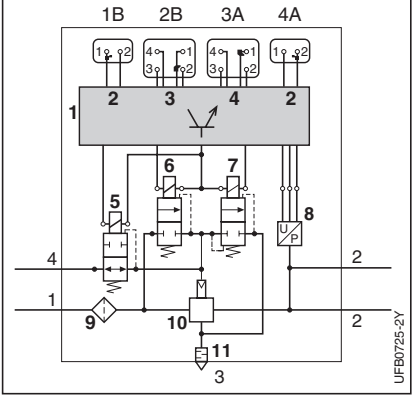
Service-brake valve
A service-brake valve for the electronic brake system is similar in design to conventional, purely pneumatic service-brake valves. However, in the service-brake valve the electronic setpoint values for brake-pressure control are also recorded (Figure 16). It thus fulfills two functions: A redundant sensor detects the driver's brake request by measuring the operating travel of the valve tappet. The measured value is transmitted to the central ECU and converted there into a braking request. In the same way as a conventional service-brake valve, the pneumatic control pressure is applied according to the actuation travel. These control pressures are required for "backup" control in the event of a fault.

Pressure-control modules
The pressure-control modules (Electro-Pneumatic Modulator, EPM, Figure 17) are the interface between the electronic brake system and the pneumatically actuated wheel brakes. They convert the required braking pressures transmitted via the CAN bus into pneumatic pressures. This conversion operation is usually performed with an inlet and outlet solenoid combination. A pressure sensor measures the controlled brake pressure and thus facilitates brake-pressure control in the closed control loop. The electrically activated "backup" valve shuts off the pneumatic control pressures of the service-brake valve in order to permit interference-free electrical pressure control.

Figure 17: Single-channel pressure-control module
1 ECU, 2 Speed sensor,
3 Brake-lining sensor,
4 CAN, 5 "Backup" valve,
6 Inlet valve, 7 Outlet valve,
8 Pressure sensor, 9 Filter,
10 Relay valve, 11 Muffler.

Connectors:
1B Connector, speed sensor 1,
2B Connector, brake-lining sensor,
3A Connector, supply and CAN bus,
4A Connector, speed sensor 2.

Ports:
1 Energy input,
2 Brake pressure to brake cylinder,
3 Vent,
4 Back-up control input.



Mounting the pressure-control modules close to the wheels means that the electrical wires for connecting the wheel-speed sensors and the brake-lining wear sensors can be kept short. The information is prepared in the pressure-control module and transmitted via the CAN bus to the central ECU. This reduces the outlay on cabling in the vehicle.

Trailer control module
The electronic trailer control module (Trailer Control Module, TCM) enables modulation of the trailer control pressure according to the functional requirements of the electronic brake system. The limits of the electrical control ranges are defined by legal requirements. The electronically specified setpoint value is converted into a physical braking pressure by means of a solenoid arrangement similar to that in the pressure-control module. The "backup" pressure is shut off either by a "backup" solenoid or by pneumatic retention, depending on the type of design adopted.

Under all normal conditions, the trailer control module must be activated by two independent control signals. This may be two pneumatic signals from two control circuits, or one pneumatic and one electrical control signal.

References for Brake systems for commercial vehicles
[1] Directive 71/320/EEC: Council Directive of 26 July 1971 on the approximation of the laws of the Member States relating to the brake systems of certain categories of motor vehicles and their trailers.
[2] ECE-R13: Uniform provisions concerning the approval of category M, N and O vehicles with regard to brakes.
[3] ISO 7638: Road vehicles – Connectors for the electrical connection of towing and towed vehicles.
[4] StVZO (road traffic licensing regulations, Germany) §41: Bremsen und Unterlegkeile (Brakes and wheel chocks).
[5] StVZO §72: Übergangsbestimmungen (Transition provisions).
[6] E. Hoepke, S. Breuer (Editors): Nutzfahrzeugtechnik. 8th Ed., Verlag Springer Vieweg, 2016.

Continuous-operation brake systems

Commercial vehicles essentially use two types of continuous-operation brake systems, deployed separately or in combination: the exhaust-brake system and the retarder.

Exhaust-brake systems
The resistance which an engine brings to the speed imposed from the outside without a fuel supply is termed the engine or exhaust brake, or drag power. The drag power of standard engines is 5 to 7 kW per liter displacement. With a pure exhaust brake the regulation according to §41 or the road traffic licensing regulations in Germany (StVZO), s.. 15 cannot be observed. Further measures can be used to increase the effect of the exhaust brake.

Exhaust-brake system with exhaust flap
In the exhaust brake with exhaust flap, a valve with a flap closes the exhaust train. The fuel supply is interrupted at the same time. As a result, back pressure is generated in the exhaust-gas system and must be overcome by each piston during its exhaust stroke (Figure 18). The braking power can be regulated by means of a pressure-control valve in the exhaust train. Additionally, this valve ensures that at high revs an excessively high pressure does not result in valve or valve-gear damage.

The exhaust brake is the most common variant used in trucks and buses, delivering a braking power of 14 to 20 kW per liter displacement.

Exhaust-brake system with constant throttle
The exhaust brake with constant throttle is also known as a decompression brake. In this system, the work performed by the engine in the compression phase is not utilized. The exhaust valves or an additional valve (constant throttle, Figure 19) is/are specifically opened at the end of the compression cycle, thereby relieving the pressure built up in the compression phase. Thus, no further work can be delivered to the crankshaft in the expansion phase.

Engine brake system with exhaust flap and constant throttle
Braking power can be further improved by a combination of exhaust flap and constant throttle (Figure 19). This combination can deliver braking power of 30 to 40 kW per liter displacement.

Figure 18: Exhaust brake with exhaust flap and additional pressure-control valve
1 Exhaust-flap actuation (compressed air),
2 Exhaust flap,
3 Bypass,
4 Pressure-control valve,
5 Exhaust,
6 Intake,
7 Piston (4th power stroke, exhaust cycle).

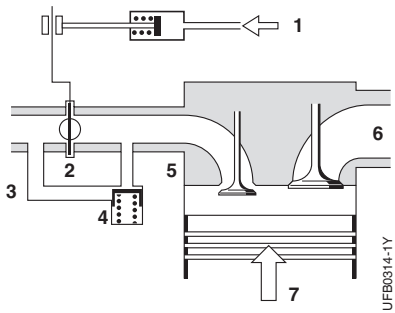
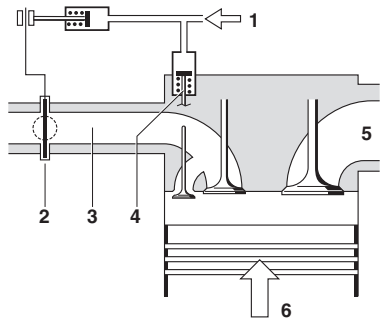


Figure 19: Exhaust brake with exhaust flap and constant throttle
1 Compressed air,
2 Exhaust flap,
3 Exhaust,
4 Constant throttle,
5 Intake,
6 Piston (2nd power stroke, compression cycle).



Retarder
Retarders are wear-free continuous-operation brakes. There are two types that differ in how they operate: hydrodynamic and electrodynamic retarders. Both systems, and the exhaust-brake systems, relieve the load on the service brake system and thereby increase the economic efficiency of the vehicle. The use of a hydrodynamic retarder can increase the service life of the service brake by a factor of 4 to 5.

In modern vehicles, retarders are incorporated into the brake-management system. The exhaust brake and the retarder are often combined as a continuous-operation brake in a vehicle. The brakes must then be activated by means of the electronic brake-management system.

Hydrodynamic retarder
Hydrodynamic retarders, also known simply as hydraulic retarders, can be subdivided into the categories of primary retarders and secondary retarders.

The primary retarder is located between the engine and the transmission, the secondary retarder between the transmission and the powered axle. Both primary and secondary retarders operate in the same way. When the retarder is activated, oil is pumped into the working area. The driven rotor accelerates this oil and transfers it at the outside diameter to the stator (Figure 20). There the oil strikes the static stator blades and is decelerated. The oil flows at the inside diameter to the rotor. The rotor's rotary motion is inhibited, and the vehicle is decelerated.

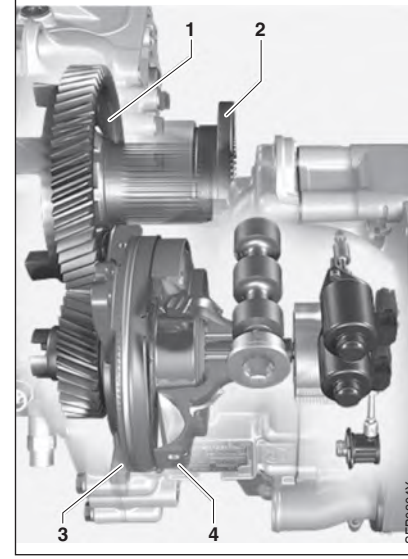
The kinetic energy is primarily converted into heat. For this reason, some of the oil must be permanently cooled by a heat exchanger.

The braking torque can be input using a hand lever or the brake pedal (in the case of a retarder integrated into the EBS electronically controlled brake system). A retarder's braking torque is dependent on the degree of fill in the working area between the rotor and the stator. The degree of fill is regulated by an ECU via a control pressure which is adjusted by proportioning valves.

A retarder can be activated hydraulically or pneumatically, whereby the braking torque can be achieved in discrete braking stages and also steplessly. Oil is primarily used as the working medium in a retarder. Current hydraulic retarders can deliver braking power up to 600 kW for brief periods. However, the continuous braking power of a retarder is dependent on the cooling capacity of the vehicle's cooling system. Modern vehicles can dissipate a continuous braking power from a retarder of 300 to 350 kW via the cooling system. Sensors are used to record overheating of the retarder or the cooling system, and if necessary the braking power is reduced under controlled conditions until the braking power equals the dissipatable quantity of heat.

Primary retarder
In the case of a primary retarder located in the drivetrain between the engine and transmission after the converter, force is

Figure 20: Functioning principle of a retarder as illustrated by a ZF intarder by way of example
1 High-speed stage,
2 Output flange,
3 Stator, 4 Rotor.



transmitted through the powered axles and the transmission in such a way that the total overrun torque is directed through the transmission. The braking effect of the primary retarder is dependent on the engine speed and the selected gear, but is not dependent on the vehicle's output speed and driving speed. This lack of dependence on the output speed is one of the major advantages of primary retarders. These are highly effective at speeds below 25 to 30 km/h (Figure 21). This is the reason why primary retarders are primarily used in vehicles which are driven at lower average speeds such as city buses and municipal vehicles. Their compact design is another advantage. A disadvantage of primary retarders is that the braking force is interrupted during a gear change. The braking force must be reduced during gear changing.

Figure 21: Operating ranges of primary and secondary retarders
1 Primary retarder, 2 Secondary retarder.

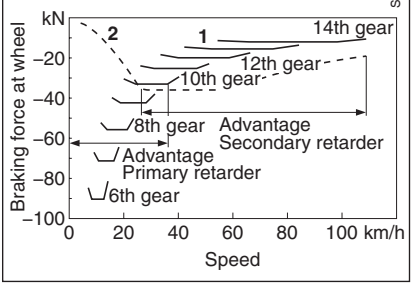
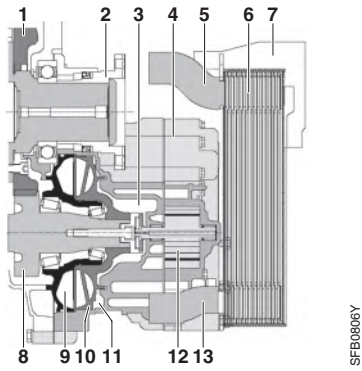


Figure 22: Hydrodynamic secondary retarder up to 600 kW braking power (ZF)
1 High-speed gear, 2 Output flange,
3 Intake passage, 4 Control housing,
5 Coolant inlet, 6 Heat exchanger,
7 Electronics, 8 Pinion shaft, 9 Stator,
10 Rotor, 11 Discharge passage,
12 Pump, 13 Coolant outlet.



Secondary retarder

In the case of a secondary retarder (Figure 22), which is located after the engine, clutch and transmission, force is transmitted directly via the output shaft. Unlike a primary retarder, there is no interruption of the braking force during a gear change with a secondary retarder. The braking effect is dependent on the ratio of the output shaft and on the driving speed. It is not dependent on the selected gear. The braking torque of a secondary retarder is very much dependent on the rotor speed. For this reason, the rotor speed is often increased by means of a high-speed stage.

The secondary retarder demonstrates great efficiency at speeds over 40 km/h (Figure 21); at speeds below 30 km/h the braking torque falls off dramatically. Because of its design, the secondary retarder can also be subsequently adapted to a transmission. The extra weight of a secondary retarder with accompanying heat exchanger and oil fill is often cited as a disadvantage, since the additional mass reduces the vehicle payload.

Secondary retarders are primarily used in long-distance vehicles which are driven at high average speeds, such as trucks and tour buses.

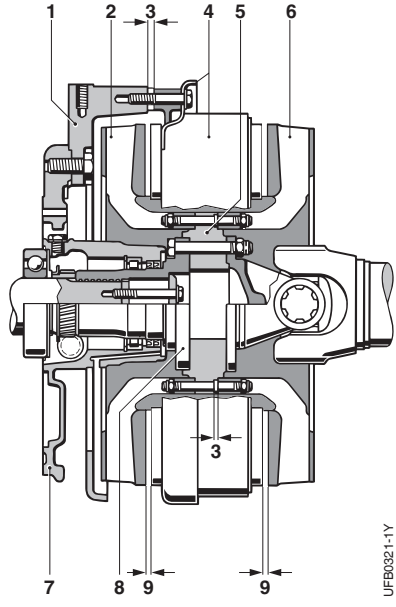
Electrodynamic retarder (eddy-current brake)

The electrodynamic retarder (Figure 23) contains two non-magnetizable steel disks (rotors) which are connected with torsional strength to the input and output shafts (here: propshaft), and a stator which is equipped with 8 or 16 coils and fastened by means of a star-shaped bracket to the vehicle frame. As soon as electric current (from the alternator or battery) flows through the coils, magnetic fields are generated which are closed by the rotors. These magnetic fields induce eddy currents in the rotors as they turn. These eddy currents in turn generate magnetic fields in the rotors which counteract the exciting

magnetic fields and thus build up a braking effect. The braking torque is determined by the strength of the excitation field, the rotational speed, and the air gap between the stator and the rotors. The braking torque decreases as the air gap increases; this air gap can be adjusted by means of spacers. Shift stages with different braking torques (Figure 24) are obtained by interconnecting the field coils in different configurations. The heat generated is dissipated by convection and radiation to atmosphere via the internally ventilated rotor disks.

As the rotors are increasingly heated, the braking power of the electrodynamic retarder decreases significantly (Figure 25). The retarder's braking power is reduced by thermal protection in order to prevent the retarder from being destroyed by excessive temperature during braking operation.

Figure 23: Electrodynamic retarder
1 Star-shaped bracket,
2 Rotor, transmission side,
3 Spacers (for adjusting air gap),
4 Stator with coils, 5 Intermediate flange,
6 Rotor, rear-axle side, 7 Transmission cover,
8 Transmission output shafts, 9 Clearance gap.



Like the primary retarder, the electrodynamic retarder is distinguished by high braking power at low engine speeds and relative design simplicity. On the downside, however, it can weigh up to 350 kg, depending on its size.

References for continuous-operation brake systems
[1] E. Hoepke, S. Breuer (Editors): Nutzfahrzeugtechnik, 8th Edition, Verlag Springer Vieweg, 2016.

Figure 24: Braking-torque characteristic of an electrodynamic retarder
4a Braking power when the cooling power limit has been reached (switching stage 4).

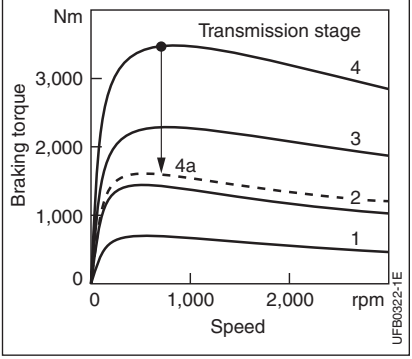
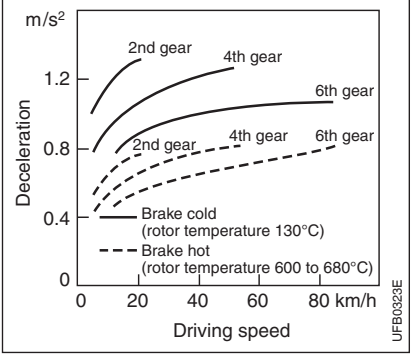


Figure 25: Influence of transmission ratio and rotor temperature on the performance of electrodynamic retarders
17 t commercial vehicle, laden.



Wheel brakes

Wheel brakes are friction brakes that convert kinetic energy into heat energy during braking. Disk and drum brakes are used as wheel brakes. Hydraulic pressure (for passenger cars) and pneumatic pressure and spring force (spring-loaded brake, for heavy commercial vehicles) are converted into an application force to press the brake pads and linings against the brake disks and drums respectively.

In passenger-car applications the thermal demands placed on wheel brakes can, in view of ever-increasing vehicle weights and higher attainable driving speeds, only be satisfied by disk brakes. Drum brakes are now only used in subcompact-size cars on the rear axle.

In the commercial-vehicle sector disk brakes have gained acceptance for on-road applications in Europe and increasingly in North America as well. In markets with less well developed road infrastructures and in applications with greater off-road usage drum brakes still play a significant role since they react more robustly to dirt and are easier to handle (maintenance, repair, etc.).

Disk brakes for passenger cars

Functioning principle

Disk brakes generate the braking forces on the surface of a brake disk that rotates with the wheel (Figure 2). The U-shaped brake caliper with the brake pads is mounted to non-rotating vehicle compo-

Figure 2: Disk brakes (diagram)
a) Fixed-caliper brake,
b) Floating-caliper brake.
1 Brake pads,
2 Piston,
3 Brake disk,
4 Brake-caliper housing,
5 Brake anchor plate.

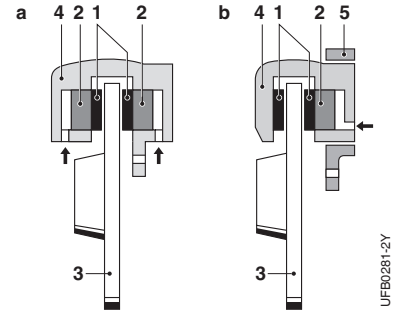
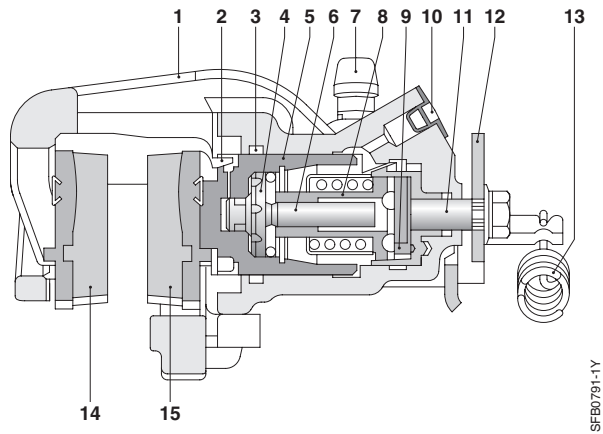


Figure 1: Floating-caliper brake with parking-brake mechanism
1 Brake-caliper housing,
2 Dust-protection seal,
3 Sealing ring,
4 Coupling,
5 Piston,
6 Threaded spindle,
7 Bleeder valve,
8 Parking-brake mechanism,
9 Cam plate,
10 Hydraulic port,
11 Shaft,
12 Parking-brake lever,
13 Spring,
14 Outer brake pad,
15 Piston-side brake pad.



nents (wheel carrier). The floating-caliper design with or without parking-brake mechanism has proven successful.

Principle of the fixed-caliper brake
In a fixed-caliper brake, both halves of the housing (flange and cover parts) are joined by the housing connecting bolts. Each half of the housing contains a piston to press the brake pad against the brake disk (Figure 2a). Ports in the housing halves connect the two pistons hydraulically.

Principle of the floating-caliper brake
In a floating-caliper brake, a piston presses the piston-side (inner) brake pad against the brake disk (Figure 2b). The generated reaction force moves the brake-caliper housing and thereby presses the outer brake pad indirectly against the brake disk. In this brake caliper the piston is therefore only seated on the inner side.

Floating-caliper brake for passenger cars
The brake caliper can be moved axially in the brake anchor plate and is guided by two sealed guide pins in the brake anchor plate (Figure 1).

Braking with the service brake
The hydraulic pressure generated by the brake master cylinder enters the cylinder chamber behind the piston via the hydraulic connection. The piston is shifted forwards and the brake pad on the piston side is applied to the brake disk. The reaction force that arises shifts the brake-caliper housing mounted in a bolt guide against the direction of piston movement. This also means that the outer brake pad is applied to the brake disk. The path of the brake pads and of the piston covered up to that point is referred to as clearance. Another increase in pressure increases the downforce of the brake pads.

Releasing the service brake
When the piston moves through the clearance, the sealing ring, which is rectangle in its initial position, is deformed. The deformed sealing ring pulls the piston back by the clearance when the hydraulic pressure drops (roll-back effect).

Braking with the integrated parking brake
When the integrated parking brake is operated, the force is transferred via the handbrake cable to the parking-brake lever. This is then twisted and the rotary motion is transferred via the shaft to the cam plate. As the balls run onto the cam plate, the piston is shifted via the pressure sleeve in the parking-brake mechanism; the threaded spindle bolted in this mechanism is shifted towards the brake pad. After crossing the clearance, first the brake pad on the piston side and then the outer brake pad are pressed against the brake disk.

Releasing the parking brake
After releasing the handbrake lever, the parking-brake lever, the shaft and the cam plate turn back to their initial positions. The pressure sleeve, the threaded spindle and the piston are pressed back into their initial position by the springs in the parking-brake mechanism. The final clearance is reached as the sealing ring reassumes its shape.

Automatic self-adjusting mechanism
Wear on the brake pads and brake disks increases the clearance and thus has to be balanced out. This automatic clearance compensation takes place during braking. The inside diameter of the rectangular piston sealing ring is slightly smaller than the piston diameter. The sealing ring thus surrounds the piston with a pre-tension. During braking the piston moves towards the brake disk and tensions the sealing ring, which as a result of its static friction can then slip on the piston only when the piston travel between brake pad and brake disk has in response to abrasion on the brake pads has become greater than the envisaged clearance. When the brake is released the piston is pulled back only by the envisaged clearance. In this way, stepless readjustment to a constant clearance is possible.

The clearance compensation of the parking-brake mechanism also takes place on application of the service brake.

The clearance of a brake caliper is approx. 0.15 mm and is thus in the range of the maximum permissible static disk run-out (axial movement per brake-disk

rotation on account of manufacturing tolerances or bearing clearances).

Brake disks

The energy converted during braking into heat is mainly absorbed by the brake disk and then dissipated to the ambient air. A fundamental distinction is made between solid disks and ventilated brake disks (Figure 3).

There are two different types of ventilated brake disk: Internally and externally ventilated. On account of the higher thermal load ventilated brake disks are used on the front axle and in high-performance or heavy vehicles also on the rear axle. Perforated or grooved friction rings improve the cooling effect and also improve wet starting response. Ventilated brake disks with radial cooling slots are used to further improve heat dissipation (Figure 3).

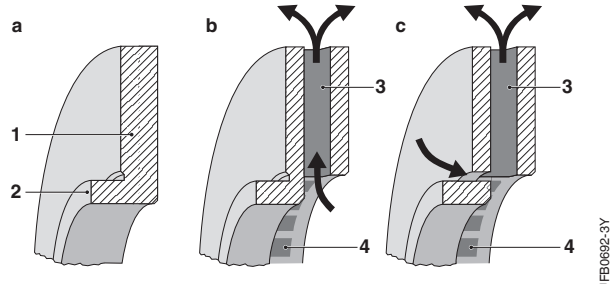
Brake disks are usually manufactured from gray cast iron. A higher proportion of carbon improves the heat-absorption capability and noise-damping performance. Alloys containing chrome or molybdenum for example also increase wear resistance. More thermally resistant brake disks made of ceramic containing silicon carbide reinforced with carbon fiber are also used on sports cars and luxury-class cars.

High corrosion requirements are imposed on brake disks due to their installation position directly on the wheel and the high mileage. Brake disks are partially or completely coated to improve corrosion resistance. High-heat-resisting paints or for example coatings containing zinc are used.

The last few years have also seen the introduction of built-up brake disks. These brake disks have a separation between the friction ring and the chamber area. The chamber area is usually made of a different material (aluminum or sheet steel) in order to achieve weight savings. These brake disks furthermore offer benefits with regard to warping (friction-ring deformation under thermal influence). A distinction is made between screwed, cast, riveted, and combined positive and non-positive connections.

Figure 3: Schematic diagram of brake disks

a) Solid brake disk,
b) Internally ventilated brake disk,
c) Externally ventilated brake disk.
1 Friction ring,
2 Chamber,
3 Cooling channel,
4 Radial cooling slots (cooling fins).



Brake pads for disk brakes
Function and requirements
The decelerating force during the braking operation is generated by sliding friction between the brake pad and the brake disk. The coefficient of friction denotes the ratio between the tensioning force generated by the brake caliper and the resulting decelerating frictional force between the pad and the disk. For passenger cars it ranges between 0.3 and 0.5 and is included in the chassis-system configuration.

As part of a safety system this function must be performed with absolute reliability and ensue in a predictable way as transparently as possible for the vehicle driver. This therefore requires that the coefficient of friction of the brake pad against the brake disk used always be identical wherever possible under all the varying – even extreme – operating conditions in wide ranges. Furthermore, the wear on the brake pads and the brake disk must be appropriately low. The dust emissions associated with wear should also be kept as low as possible and the composition of these emissions should pollute the environment as little as possible. Brake systems should also operate with as little noise as possible.

Optimum tuning of the brake system to the vehicle is quite significantly achieved via the properties of the brake pad, the design and manufacture of which, of all the components in the system, opens up the most wide-ranging possibilities.

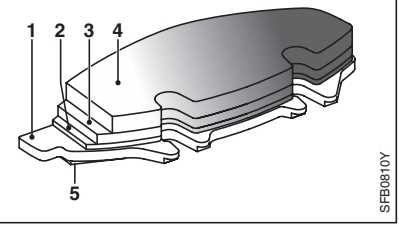
Design
The brake pad (Figure 4) consists of a number of parts whose properties must all be meticulously matched to each other to meet the complex functional requirements. A back plate (carrier) usually made of steel serves to support the braking force exerted by the friction lining against the brake caliper and conversely to transfer the piston force as uniformly as possible to the lining friction surface. To this end, the back plate must adhere to exact dimensional tolerances so as to avoid braking noises or unreliable functioning. In addition, the material must exhibit sufficient strength so as not to permit any plastic deformation at the contact faces or hooks. In many a design the back plate still supports the strength of the brake caliper.

For the most part, anti-noise shims or films are glued, riveted or even attached with clips to the back of the brake pad painted for corrosion-protection purposes. They are for the most part essential for avoiding noise emissions. It is important here to ensure that the shims cannot be appreciably displaced over the period of operation. Otherwise, unwanted contact between the anti-noise shim and brake or wheel components may result under adverse circumstances. Clips for positioning and locating in the piston and caliper may still be required for fitting in the brake.

Manufacture
The most cost-effective solution is, after cleaning off oxide coatings and oil residues, to apply a coating of a highly thermally and chemically stable adhesive. Racing cars and a number of heavy, high-performance premium vehicles also use bonding variants which establish positive locking between the pad and the carrier, and thus are above all much more thermally resistant but at the same time significantly more expensive than an adhesive bonding.

Finally, the friction lining is applied. The friction lining is for the most part pressed from a molding powder or granulate in a heated press onto the back plate. At temperatures between 130 °C and around 170 °C the binder component of the friction-lining mixture combines permanently

Figure 4: Structure of a brake pad for disk brakes
1 Back plate,
2 Adhesive layer,
3 Intermediate layer,
4 Friction lining,
5 Anti-noise shim.



with the back-plate adhesive to form a component. In the process all the other ingredients are also integrated into the binder matrix usually consisting of highly chemically and thermally resistant phenolic resin. This so-called hot-pressing process with dry mixtures has become established in the mass market as the best combination of reliability, range of adjustments of physical properties, product variance, and costs.

Composition of brake pads

The recipe for the molding compound of a typical European brake pad consists of up to more than twenty different ingredients. These can be grouped into the categories of binders, metals, graphite and coke, fillers, organic fibers, abrasives, and lubricants. Table 1 shows by way of example some of the raw materials normally used in brake pads. Aside from their concentration in the recipe, their effect in the brake pad is also determined by their individual composition, microstructure, and particle size. Reliable braking performance with low wear is obtained only by choosing ideal ingredients and carefully optimizing the concentrations of the metal, abrasive, and lubricant proportions. Physical-mechanical properties for sufficient strength and particularly also for comfortable, noise-free braking are adjusted by means of tribologically neutral fillers in broad concentration ranges. Only with this complexity of recipes can the high, sometimes contradictory requirements regarding friction loss, reliability, wear, and braking comfort be optimally reconciled for the majority of drivers.

Between the friction lining and the adhesive a roughly 2…3 mm thick intermediate layer similar mechanically and chemically to the friction lining can be used for example to improve the pad connection or damp braking noises.

Finally, a brake pad must be clearly marked to denote its country of use in order for example to display the regulatory approval for use in public traffic. The markings are usually printed or stamped on the backs of the brake pads.

Special regional requirements

Unlike in Europe – and in particular unlike in Germany – in many regions around the world drivers cannot drive very fast by vehicle since speed limits are enforced even on expressways and freeways. Resulting lower friction-coefficient requirements, but higher requirements with regard to wear and dust formation led for example in Asia and the USA to the development of different friction materials from those in Europe. From Japan come the so-called NAO materials (Non Asbestos Organics), which are characterized by very low brake-disk corrosion and consequently also by low dust formation. However, at present these recipes cannot deliver friction coefficients that are as high as those produced by the recipes customary in Europe. In contrast to European recipes, they typically contain no steel wool or iron powder. They are therefore also referred to as non-steel pads or more promotionally effectively as ceramics. In contrast, the materials customary in Europe are called low-steel or low-met pads. The combination of the good friction loss of European recipes with the low dust generation of NAO rec-

Table 1: Raw-material groups with typical concentration ranges of brake-pad recipes for use in Europe

Raw-material group	Raw material	Volume %
Metals	Steel wool Aluminum wool Copper wool Zinc powder	10…15
Binders	Phenolic resin Caoutchouc	15…20
Fillers	Silicates, e.g. Mica powder Talcum Chalk	20…50
Abrasives	Aluminum oxide Silicon carbide	2…5
Lubricants	Molybdenum sulfide Tin sulfide	2…10
Organic fibers	Aramide fiber Cellulose fiber	2…5
Graphite and coke	Graphite and coke	10…25

ipes constitute a current conflict of goals for the development of friction materials.

Up to half the weight of semi-met pads is made up of iron materials. They represent a good cost compromise with weaknesses in the high-load and high-temperature ranges.

Disk brakes for commercial vehicles

Special disk brakes have been developed for commercial vehicles. These disk brakes are actuated with compressed air. Because the pressure here is much smaller than that in a hydraulic brake, the brake cylinders cannot be integrated into the brake calipers. They must be flanged into position (Figure 5).

Functioning principle

Operating concept of the service brake

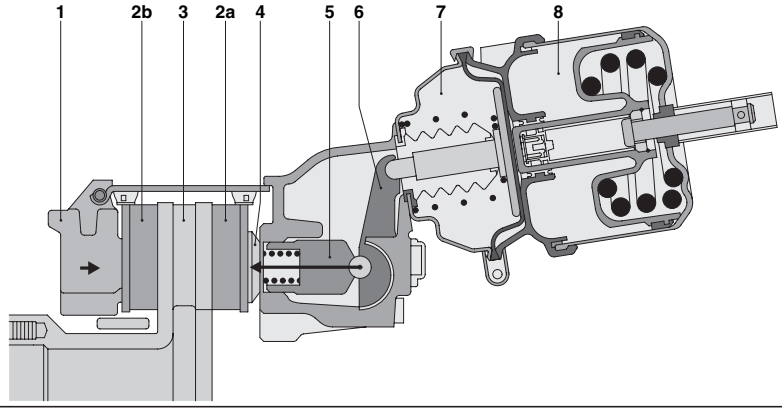
When air enters the service brake cylinder, the eccentrically mounted brake lever is actuated. The brake cylinder force is boosted by the lever ratio and transferred via the bridge and plungers to the inner brake pad. The reaction force that arises at the brake caliper is transferred by shifting the brake caliper to the outer brake pad.

Operating concept of the parking brake

When the spring-type brake cylinder is vented, the force of the pre-tensioning springs is released. Via the spring-type brake piston, this moves the piston and the push rod of the service brake cylinder to operate the brake. In the case of the parking brake, the pressure in the spring-type brake actuator is completely

Figure 5: Disk brake with combination brake cylinder

1 Brake caliper, 2a Inner brake pad, 2b Outer brake pad, 3 Brake disk, 4 Plunger,
5 Bridge, 6 Eccentrically mounted brake lever, 7 Service-brake cylinder, 8 Spring-type actuator.



eliminated, thus releasing the force of the pre-tensioning springs to achieve maximum braking effect.

Automatic self-adjusting mechanism
Disk brakes pneumatically or mechanically actuated by the spring-type actuator are fitted with automatic clearance compensation.

Wear monitoring
Continuous wear monitoring may also be provided. This is required in the case of electronic brake systems for wear adaptation and for service information systems.

Brake disks
Solid brake disks are used less frequently on commercial vehicles because they can only dissipate the heat slowly. Internally ventilated brake disks have a larger surface through which heat exchange can take place. In this design two friction rings are connected via bridges. The rotation of the brake disk creates on the inside a radial ventilation effect in the outward direction.

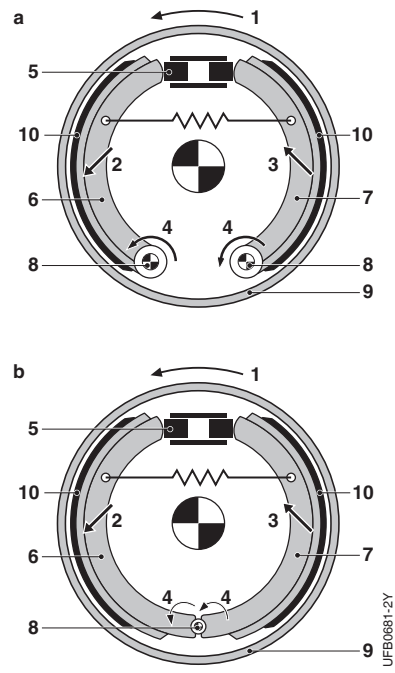
Brake disks for commercial vehicles are usually manufactured from gray cast iron. The carbon content increased up to the saturation limit provides good thermal conductivity.

Drum brakes

Drum brakes are radial brakes with two brake shoes. They generate their braking force on the inner friction surface of a brake drum.

Drum-brake designs
There are two different designs of drum brake, based on how the brake shoes are guided:
– brake shoes with fixed pivot (Figures 6a and 6b),
– brake shoes as sliding shoes (Figure 7).

Figure 6: Principle of simplex brake
a) Brake shoe with two single pivots,
b) Brake shoe with one double pivot.
1 Direction of rotation of brake drum with vehicle moving forwards,
2 Self-augmentation, 3 Self-inhibition,
4 Torque,
5 Double-acting wheel-brake cylinder,
6 Leading brake shoe (primary shoe),
7 Trailing brake shoe (secondary shoe),
8 Fulcrum (pivot),
9 Brake drum, 10 Brake lining.

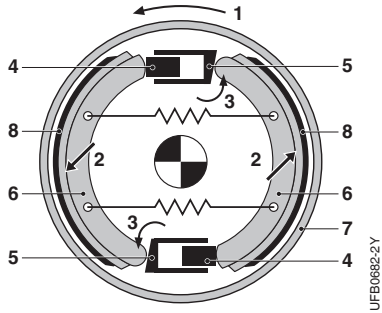


At the brake shoe that rotates in the direction of the brake drum (primary shoe, leading brake shoe, Figure 6), the friction force during the braking operation creates a turning force around the brake-shoe fulcrum which in addition to the application force presses the shoe against the drum. This generates a self-augmenting effect. In a simplex brake a turning force is created around the fulcrum of the trailing brake shoe (secondary shoe) which diminishes the applied application force. This therefore creates a self-inhibiting effect.

Sliding-shoe guides are used in simplex, duplex, duo-duplex, servo and duo-servo brakes. Brake shoes with a fixed pivot are subject to unequal levels of wear in that they cannot center themselves like sliding shoes.

Principle of the simplex brake
A double-acting wheel-brake cylinder actuates the brake shoes (Figures 6a and 6b). The fulcrums of the brake shoes are pivots (two single pivots or one double pivot). When the vehicle is moving forwards, self-augmentation affects the leading brake shoe and self-inhibition affects the trailing brake shoe; the pattern is the same when the vehicle is backing up.

Figure 7: Principle of duplex brake
1 Direction of rotation of brake drum with vehicle moving forwards,
2 Self-augmentation, 3 Torque,
4 Wheel-brake cylinder, 5 Fulcrums,
6 Brake shoes, 7 Brake drum,
8 Brake lining.



Principle of the duplex brake
Each brake shoe is actuated by a single-acting wheel-brake cylinder (Figure 7). The brake shoes designed as sliding shoes are supported on the back of the opposing wheel-brake cylinder. The duplex brake is single-acting, i.e., it has two leading self-augmenting brake shoes when the vehicle is moving forwards. There is no self-augmentation when the vehicle is backing up.

Simplex drum brake
Functioning principle of passenger-car brake
The principle of a drum brake is explained using a hydraulically operated simplex drum brake with integrated parking brake and automatic self-adjusting mechanism as an example (Figure 8). Other drum brake designs (e.g., duplex brake, duo-duplex brake) are rarely used today.

When driving extension springs pull the two brake shoes away from the brake drum so that a clearance is created between the drum friction surface and the brake linings. In the case of simplex brakes, a two-sided hydraulic wheel-brake cylinder generates the application force for the brake shoes during braking by converting the hydraulic pressure into mechanical force. Here, the leading and the trailing brake shoes with the brake pads press against the brake drums. The other ends of the brake shoes on the opposite side to the wheel-brake cylinder are braced by a support bearing that is attached to the brake anchor plate.

The leading brake shoe (primary shoe) generates a higher proportion of braking torque than the trailing brake shoe (secondary shoe). Wear is therefore greater on the primary lining. This lining is thicker or longer in design to compensate.

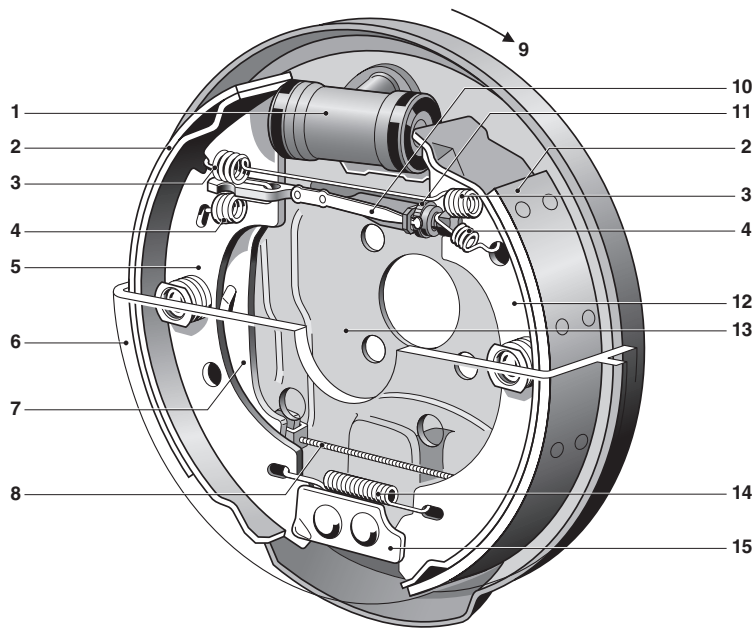
Functioning principle of simplex brake with S-cam for commercial vehicles
In commercial vehicles with compressed-air brake systems, the application force is frequently generated by a rotatable S-cam. S-cam rotation is effected by the brake cylinder, the brake lever (slack adjuster), and the brake-cam shaft (Figure 9).

Functioning principle of wedge-actuated brake
Wedge-actuated brakes are also used in commercial vehicles. Here, the application force for the brake shoes is generated by a wedge actuated by the brake cylinder (Figure 10).

During braking the diaphragm brake cylinder is exposed to compressed air. This displaces the wedge to the right. The wedge slides between the pressure rollers. These roll on the wedge and the thrust members. The generated application force is transferred via the thrust members to the brake shoes. The excessive clearance created by brake-lining wear is compensated for by the adjusting mechanism.

Automatic adjusters
Wheel brakes must be fitted with adjusters to compensate for the increased clearance caused by lining wear. The brakes must be easily adjustable or have an au-

Figure 8: Simplex drum brake with integrated parking brake
1 Wheel-brake cylinder, 2 Brake lining, 3 Extension spring (for brake shoes),
4 Extension spring (for adjuster), 5 Trailing brake shoe,
6 Brake drum, 7 Parking-brake lever, 8 Brake cable,
9 Direction of drum rotation, 10 Thermocouple (adjuster),
11 Adjuster wheel (with elbow lever), 12 Leading brake shoe,
13 Brake anchor plate, 14 Extension spring (for brake shoes),
15 Brake-shoe pin bushing.



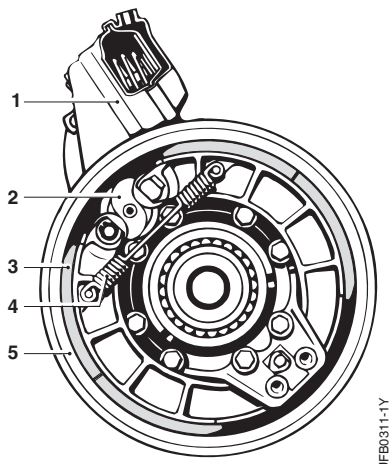
tomatic adjuster (§ 41 s. 1 StVZO [1], ECE R13-H [2]).

On simplex drum brakes for passenger cars, the adjuster is part of the push rod or pressure sleeve situated under initial spring tension between the brake shoes. When the permissible clearance is exceeded, the adjuster automatically lengthens the push rod or pressure sleeve (to different extents depending on the adjuster design) and thereby adjusts the clearance between brake shoe and brake drum. Automatic adjusters mostly operate in conjunction with a thermocouple on a temperature-sensitive basis in order to prevent adjustment when the brake drum is hot (expanded).

In commercial vehicles with S-cams, the adjuster is part of the brake lever. For this purpose it has manual adjustment or is designed as an automatic linkage adjuster.

In wedge-actuated brakes, an automatically acting adjuster is integrated into the wedge mechanism.

Figure 9: Simplex drum brake with S-cam
1 Diaphragm-type cylinder,
2 S-cam,
3 Brake shoes,
4 Return spring,
5 Brake drum.



Parking brake

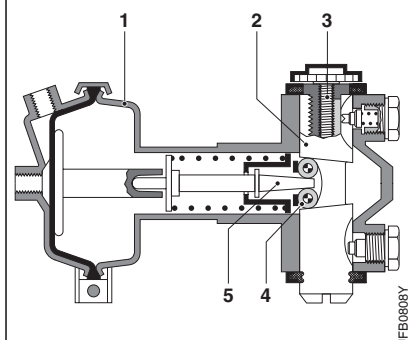
A parking brake is integrated in the drum brake. The brake cable is actuated either via the handbrake lever inside the vehicle or via an electric motor with spindle. The handbrake lever is mounted at the top of the trailing brake shoe. When the parking brake is operated, the brake cable pulls the handbrake lever downwards to the right, causing the handbrake lever to press the brake shoes via the push rod against the brake drum.

In a commercial vehicle the drum brake is actuated as a parking brake by venting the spring-type brake cylinder.

References

[1] §41 Straßenverkehrs-Zulassungsordnung (Road traffic licensing regulations, Germany). Bremsen und Unterlegkeile (Brakes and wheel chocks). [2] ECE R13-H: Regulation No. 13-H of the United Nations Economic Commission for Europe (UN/ECE) – Uniform provisions concerning the approval of passenger cars with regard to braking.
[3] B. Breuer, K. Bill (Editors): Bremsenhandbuch: Grundlagen, Komponenten, Systeme, Fahrdynamik. 5th Ed., Verlag Springer Vieweg, 2017.

Figure 10: Wedge-actuated brake
1 Diaphragm brake cylinder,
2 Thrust member,
3 Adjusting mechanism,
4 Pressure rollers,
5 Wedge.



Wheel-slip control systems

Function and requirements

Slip

When starting off, accelerating and braking, the efficiency required to transfer forces to the road depends on the traction available between the tires and the road surface. Slip occurs when the speed v_R at which the wheel center moves in the longitudinal direction (vehicle speed) differs from the speed v_U at which circumference rolls.Brake slip λ_B and drive slip λ_A are calculated as follows:

$$\lambda_B = \frac{v_U - v_R}{v_R}\,,\ \lambda_A = \frac{v_U - v_R}{v_U}\,.$$

When a wheel is locked, according to this definition brake slip $\lambda_B = -1$; when the vehicle is stationary with spinning wheels drive slip $\lambda_A = 1$ (see Slip, Fundamentals of automotive engineering).

Adhesion/slip curves

The tire must roll with slip so that it can transmit force to the road. A tire without slip would not deform on the wheel contact area and therefore could not transmit either a longitudinal force or a lateral force. The transmittable forces are dependent on the slip. The adhesion/slip curves (Figure 2) illustrate this dependence. They progress identically for braking and propulsion/traction.

Wheel-slip control systems ensure the optimum transmission of forces between tires and road to keep the vehicle directionally stable and more easily controllable for the driver. For this purpose the longitudinal slip of the individual wheels – i.e.,

Figure 2: Adhesion/slip curve
Curve shape for dry road surface, $\mu_{HF} \approx 0.8$.
1 Braking or tractive force,
2 Lateral force.

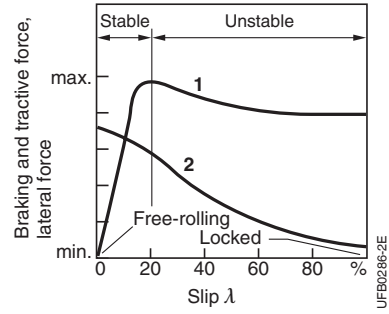
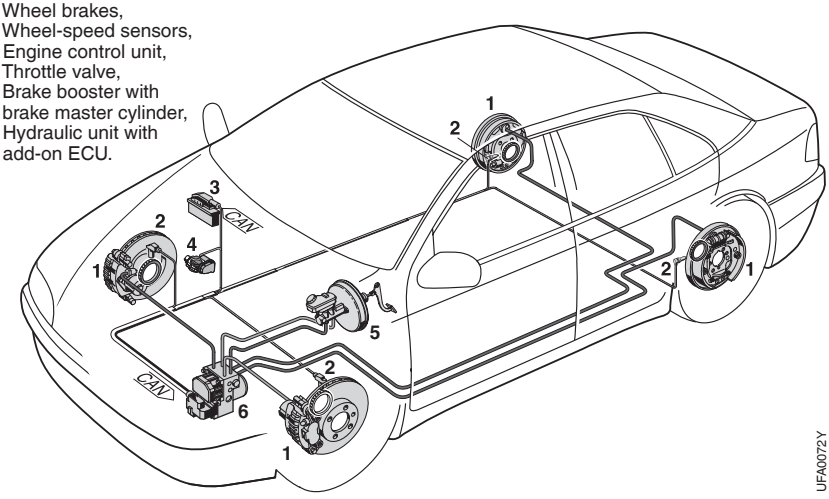


Figure 1: System diagram of a brake system with ABS
1 Wheel brakes,
2 Wheel-speed sensors,
3 Engine control unit,
4 Throttle valve,
5 Brake booster with brake master cylinder,
6 Hydraulic unit with add-on ECU.



the wheel speed referred to the velocity of the wheel center in the longitudinal direction – is regulated by modulating the braking or drive torque. There is a fundamental distinction between the antilock braking function (ABS, antilock braking system) and the traction control system (TCS).

The vast majority of acceleration and braking operations involve only limited amounts of slip, allowing response to remain within the stable range in the adhesion/slip curves. Any rise in slip (brake slip when braking and drive slip when accelerating) is accompanied initially by a corresponding increase in adhesion. Beyond this point, any further increases in slip take the curves through the maxima and into the unstable range (Figure 2) where any further increase in slip generally results in a reduction in adhesion. When braking, this results in wheel lock within a few tenths of a second. When accelerating, one or both of the driven wheels start to spin more and more as the drive torque exceeds the adhesion by an ever increasing amount.

Effect of ABS and TCS

As brake slip increases the ABS function becomes active and prevents the wheels from locking; as drive slip increases TCS prevents the wheels from spinning. Thanks to ABS the vehicle retains its directional stability and steerability even under emergency braking on a slippery road surface. The dangerous phenomenon of jackknifing is also prevented in commercial-vehicle combinations. The TCS function optimizes the transmission of forces of the drive wheels when accelerating and thereby improves both traction and stability.

Figure 1 shows an ABS system with its components for a passenger car with a hydraulic brake system. In contrast to passenger cars, commercial vehicles have pneumatic power-brake systems (air brakes). Nevertheless, the functional description of an ABS or TCS control process for passenger cars also applies in principle to commercial vehicles.

The functions of ABS and TCS have in the meantime been integrated into the driving-dynamics control.

Control systems

ABS control

Basic closed-loop control process

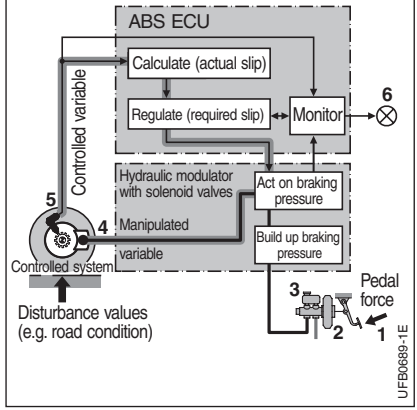
The wheel-speed sensor senses the state of motion of the wheel (Figure 3). If one of the wheels shows signs of incipient lock, there is a sharp rise in peripheral wheel deceleration and in wheel slip. If these exceed defined critical levels, the ABS controller sends commands to the solenoid-valve unit (hydraulic unit) to stop increasing or to reduce wheel brake pressure until the danger of wheel lock is averted. The braking pressure must then rise again to ensure that the wheel is not underbraked. During automatic brake control, the stability or instability of wheel motion must be detected constantly, and kept within the slip range at maximum braking force by a sequence of pressure-rise, pressure-retention and pressure-drop phases.

Typical control cycle

The control cycle depicted in Figure 4 shows automatic brake control in the case of a high friction coefficient. The change in wheel speed (braking deceleration) is calculated in the ECU. After the value

Figure 3: ABS control loop of a passenger-car system

1 Brake pedal, 2 Brake booster,
3 Brake master cylinder with fluid reservoir,
4 Wheel-brake cylinder,
5 Wheel-speed sensor,
6 Indicator lamp.



falls below the (–a) threshold, the hydraulic-unit valve unit is switched to pressure-holding mode. If the wheel speed then also drops below the slip-switching threshold λ_1, the valve unit is switched to pressure drop; this continues as long as the (–a) signal is applied. During the following pressure-holding phase, peripheral wheel acceleration increases until the (+a) threshold is exceeded; the braking pressure is then kept at a constant level.

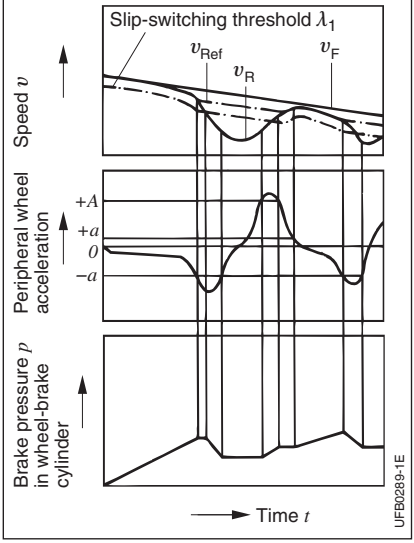
After the relatively high (+A) threshold has been exceeded, the braking pressure is increased, so that the wheel is not accelerating excessively as it enters the stable range of the adhesion/slip curve. After the (+a) signal has dropped off, the braking pressure is slowly raised until, when the wheel acceleration again falls below the (–a) threshold, the second control cycle is initiated, this time with an immediate pressure drop.

In the first control cycle, a short pressure-holding phase was initially necessary to filter out any faults. In the case of high wheel moments of inertia, low friction coefficient and slow pressure rise in the wheel-brake cylinder (cautious initial braking, e.g., on black ice), the wheel might lock without any response from the deceleration switching threshold. In this case, therefore, the wheel slip is also included as a parameter in the brake-control system.

Under certain road-surface conditions, passenger cars with all-wheel drive and with differential locks engaged pose problems when the ABS system is in operation; this calls for special measures to take into account the reference speed during the control process, lower the wheel-deceleration thresholds, and reduce the engine-drag torque.

Figure 4: ABS control cycle for high friction coefficients

v_{Ref} Reference speed,
v_U Wheel circumferential speed,
v_F Vehicle speed,
a,A Wheel-deceleration thresholds.



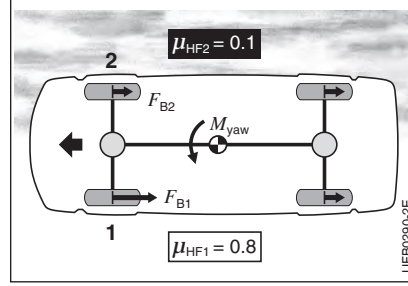
Control cycle with yaw-moment build-up delay

When the brakes are applied on a road surface with uneven grip (μ split: left-hand wheels on dry asphalt, right-hand wheels on ice), vastly different braking forces at the front wheels result and induce a turning force (yaw moment) about the vehicle's vertical axis (Figure 5).

On smaller cars, ABS must be supplemented by an additional yaw-moment build-up delay device to ensure that control is maintained during panic braking on asymmetrical road surfaces. Yaw-moment build-up delay holds back the pressure rise in the wheel-brake cylinder on the

Figure 5: Yaw-moment build-up induced by large differences in friction coefficients

M_{yaw} Yaw moment, F_B Braking force,
μ_{HF} Coefficient of friction.
1 "High wheel", 2 "Low wheel".



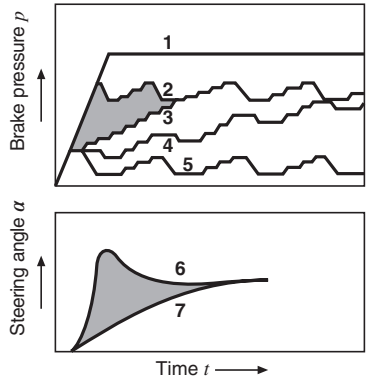
front wheel with the higher coefficient of friction at the road surface ("high wheel").

The yaw-moment build-up delay concept is demonstrated in Figure 6: Curve 1 represents the brake-master-cylinder pressure p_{MC}. Without yaw-moment build-up delay, the braking pressure at the wheel running on asphalt quickly reaches p_{high} (Curve 2), while the braking pressure at the wheel running on ice rises only to p_{low} (Curve 5); each wheel brakes with the maximum transferable braking force (see Individual control).

The yaw-moment build-up delay 1 system (Curve 3) is designed for use on vehicles with less critical handling characteristics, while yaw-moment build-up delay 2 is designed for cars which display an especially marked tendency toward yaw-induced instability (Curve 4). In all cases in which yaw-moment build-up delay comes into effect, the high wheel is under-braked at first. This means that the yaw-moment build-up delay must always be very carefully adapted to the vehicle in question in order to limit increases in stopping distances.

Figure 6: Curves for braking-pressure/ steering-angle characteristic with yaw-moment build-up delay

1 Brake-master-cylinder pressure p_{MC},
2 Brake pressure p_{high} w/o YMBD,
3 Brake pressure p_{high} w/ YMBD 1,
4 Brake pressure p_{high} w/ YMBD 2,
5 Brake pressure p_{low} at "low wheel",
6 Req. steering angle α w/o YMBD,
7 Req. steering angle α w/ YMBD.



Curve 6 in Figure 6 shows that for an ABS system without yaw-moment build-up delay a significantly higher steering angle is required when countersteering.

ABS control methods

The axle-based ABS control methods differ essentially in the number of control channels and the behavior when braking at μ split.

Individual control

Individual control, whereby each wheel is individually slip-controlled, produces the shortest braking distances. The drawback, however, is the yaw moment occurring under μ-split conditions, which must be compensated for by appropriate countersteering. This method is used exclusively on the rear axle since the steering and yaw moments occurring at the front axle would not be controllable for the driver when braking at μ split.

Select-low control

Select-low control (SL) is used to avoid yaw and steering moments entirely. Here, single-channel wheel-slip control is effected at the wheel with the lowest friction coefficient (select low), as a result of which both wheels on one axle receive the same brake pressure. Therefore only one single pressure-control channel per axle is required. Under μ-split conditions this produces optimum steerability and directional stability at the expense of braking distance. In the case of homogenous friction coefficients, braking distance, steerability, and directional stability are similar to those of the other methods.

Individual control, modified

"Individual control, modified" (IRM) has proven to be a good compromise between steerability, stability, and braking distance. This two-channel control method necessitates a pressure-control channel at each wheel on the axle. By appropriately limiting the brake-pressure difference between the right and left sides, the yaw and steering moments are restricted to a controllable extent. This results in a braking distance which is only a little longer than that for individual control, but it does ensure that vehicles with critical handling characteristics remain controllable.

TCS control
The traction control system has two fundamental tasks:
– optimizing traction by utilizing the available friction coefficient in the best possible way,
– ensuring vehicle stability (directional stability) by preventing the drive wheels from spinning.

To optimize traction all the driven wheels must utilize their individual friction coefficients to the maximum to the best possible extent. To this end, the wheel speeds are synchronized by active braking of the spinning wheels (brake controller or electronic differential-lock function). In this way, the braking torque exerted on the spinning wheel is available through transmission by the differential to the non-spinning wheel as drive torque.

To ensure directional stability, the wheel slip is controlled with the aid of the drive torque by the engine controller in such a way as to achieve the best possible compromise between traction and lateral stability.

Using the brake-control function described above, the driving wheels can also be synchronized so that a mechanical differential lock, if fitted, can be activated automatically, e.g., with the aid of a pneumatic cylinder. The ABS/TCS ECU calculates the correct point and conditions for releasing the differential lock.

In contrast to mechanical differential locks, the tires do not scrub on tight corners. A fundamental observation about this type of system (when it assumes an electronic brake-control function) is that it is not intended for continuous use on difficult offroad terrain. Since the brake-control function is achieved by braking the relevant wheel, brake heating is an inevitable consequence.

For multi-axle vehicles with complex drive configurations and a number of differentials (e.g., 6×4 or 8×6 with three or five differentials) the TCS function can control up to six wheels individually.

Engine drag-torque control
Particularly at low load with a low friction coefficient or very powerful engines the wheels on the powered axle can lock due to high engine-drag torques (e.g., when downshifting), which results in unstable handling. In this case, engine drag-torque control increases the wheel speeds by increasing the drive torque and thereby prevents the impending instability. The actively exerted drive torque is limited for safety reasons.

Electronic load-dependent braking-force regulation
Load differences and the dynamic axle-load shift that occurs during sharp braking call for an adaptation of the braking forces. This was originally performed by ALB valves (automatic load-dependent braking-force regulator), which reduce the brake pressure usually on the rear axle depending on the axle load. In current ABS systems this function is assumed by electronic load-dependent braking-force regulation. Here, under minimal deceleration conditions, the differential slip between the front and rear axles is minimized, whereby the brake pressure on the rear axle is electronically reduced. This results, assuming the same friction conditions on both axles, in identical braking and consequently optimum braking under driving-dynamics considerations. This function dispenses with the need for the additional ALB valve.

ABS/TCS systems for passenger cars

The requirements imposed on an ABS system are described in the regulations ECE-R13 [1]. This regulation defines ABS as a component of a service-brake system (Figure 1) which automatically controls wheel slip in the direction of wheel rotation on one or more wheels when braked.

ECE-R13 Annex 13 defines three categories. The present generation of ABS meets the highest level of requirements (Category 1).

Components

An ABS or ABS/TCS system consists of the following components: wheel-speed sensors, electronic control unit (ECU), and hydraulic unit (for passenger cars) or pressure-control valves (for commercial vehicles).

Wheel-speed sensors

The most important input variables for controlling wheel slip are the wheel speeds, which are recorded by wheel-speed sensors. These sensors scan a rotating sensor ring and generate an electrical signal with a frequency proportional to the rotational speed (see rpm sensors).

There are basically two different types of rpm sensor: active and passive. The active rpm sensors used predominantly in passenger cars operate according to the Hall principle and can record, aside from the road-speed signal, further information such as the temperature and transmit it to the ECU.

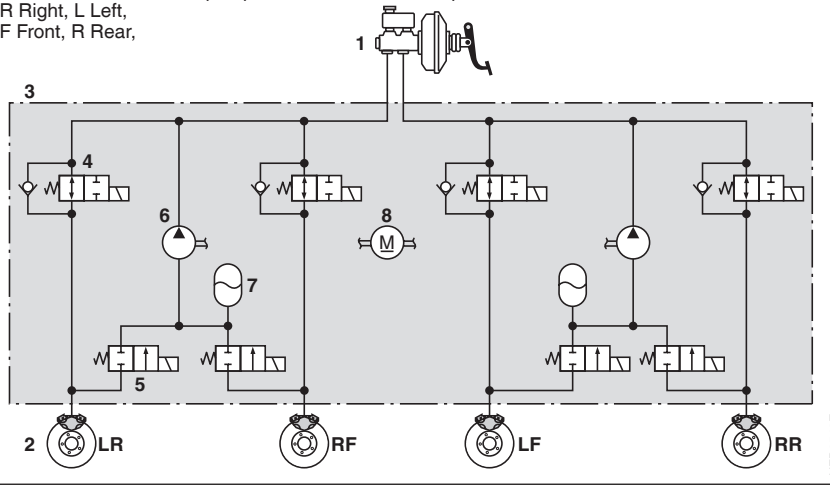
Hydraulic unit of an ABS system

The main hydraulic components of the hydraulic unit – also called the hydraulic modulator – are the following (Figure 7):

- one return pump per brake circuit,
- accumulator chamber,
- damping functions, previously performed by an accumulator chamber and a flow restrictor, are now performed both hydraulically and by control systems, i.e., software,
- 2/2-way solenoid valves with two hydraulic positions and two hydraulic connections.

There is one pair of solenoid valves for each wheel (except in the case of 3-channel configurations with front/rear brake-circuit split, see ABS system variants) – one of which is open when de-energized for pressure rise (inlet valve, IV) and one which is closed when de-en-

Figure 7: Hydraulic system of an antilock braking system.
1 Brake master cylinder, 2 Wheel-brake cylinder, 3 Hydraulic unit, 4 Inlet valve, 5 Outlet valve, 6 Return pump, 7 Accumulator, 8 Pump motor.
R Right, L Left,
F Front, R Rear,



ergized for pressure drop (outlet valve, OV). In order to achieve rapid pressure relief of the wheel brakes when the pedal is released, the inlet valves each have a non-return valve which is integrated into the valve body (e.g., non-return valve sleeves or unsprung non-return valves).

The assignment of pressure-rise and pressure-drop functions to separate solenoid valves with only one active (energized) setting has resulted in compact valve designs, i.e., smaller size and weight, as well as lower magnetic forces compared to the previous 3/3-way solenoid valves. This allows optimized electrical control with low electrical power loss in the solenoid coils and the control unit. In addition, the valve block (Figure 8) can be made smaller. This results in quite significant savings in weight and size.

The 2/2-way solenoid valves are available in a variety of designs and specifications, and, because of their compact dimensions and excellent dynamics, they allow fast electrical switching times sufficient for pulse-width-modulated cyclic operation. In other words, they have “proportional-valve characteristics”.

ABS 8 from Bosch (Figure 8) benefits from current-signal-modulated valve control which substantially improves function (e.g., adaptation to changes in coefficient of friction) and ease of control (e.g., smaller deceleration fluctuations with the aid of pressure stages and analog pressure control). This mechatronic optimization has positive effects not only on function but also on user-friendliness, i.e., noise and pedal feedback.

ABS 8 is capable of specific adaptation to individual vehicle-class requirements by varying the components (e.g., using motors of different power ratings, varying accumulator chamber size, etc.). The power of the return motor can vary within a range of approx. 90 to 200 watts. The accumulator-chamber size is also variable.

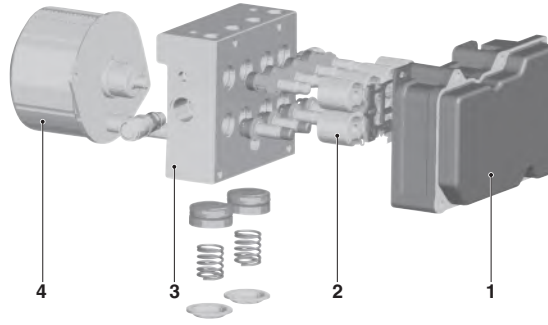
Hydraulic unit of an ABS/TCS system

On passenger cars with hydraulic brake systems, an expanded ABS hydraulic unit is required for TCS brake intervention. Depending on the variant, the expansion can comprise an intake valve and a changeover valve (Figure 9). An additional hydraulic presupply pump and a pressure accumulator may be required. During a necessary brake intervention, the intake and changeover valves assigned to the spinning wheel and the ABS return pump are electrically actuated. The return pump can draw brake fluid from the brake master cylinder through the intake valve. The changeover valve blocks the return flow to the brake master cylinder. The pressure generated by the return pump passes through the inlet valve to the wheel-brake cylinder of the spinning wheel, as a result of which the wheel is braked and prevented from spinning. Braking pressure is built up, as dictated by the situation and adapted by continuous monitoring of the control process, by alternating and electrically clocked actuation of the inlet and outlet valves in the hydraulic unit.

On completion of the control phase, electrical actuation is terminated and the braking pressure applied for TCS control,

Figure 8: Design of ABS 8 hydraulic unit from Bosch

1 ECU,
2 Coil group,
3 Hydraulic unit,
4 Pump motor.



as following a normal braking operation, is reduced via the intake and changeover valves and the brake master cylinder.

Electronic control unit (ECU)
The ECU processes the signals supplied by the wheel-speed sensors. After these signals have been conditioned and filtered, a vehicle reference speed is calculated and this forms the basis for the wheel-slip calculation. To form the reference speed the individual rotational-speed signals are, depending on the respective driving situation and other criteria, corrected if necessary with different weighting.

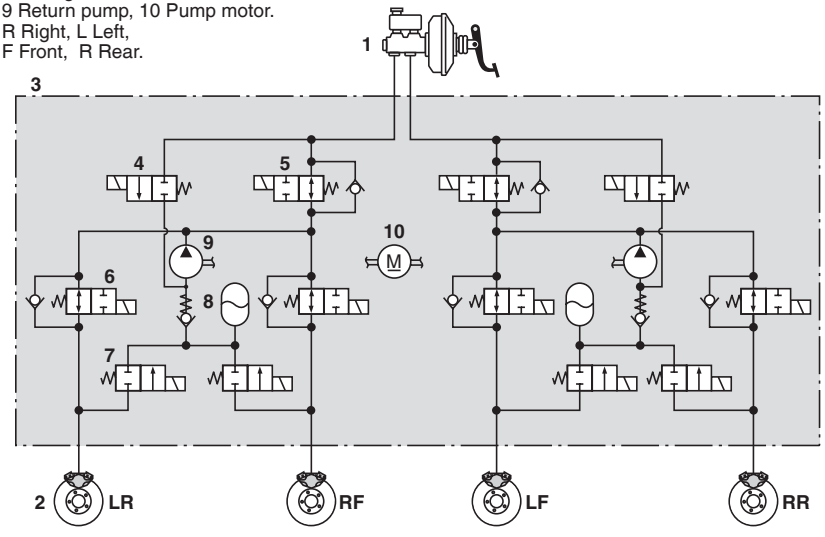
The associated solenoid valves are activated depending on the individual wheel-slip values and the target values.

Because of the safety relevance of the slip-control functions, the ECUs contain comprehensive safety and diagnostic functions to monitor the complete system on a permanent basis. Detected faults lead to partial or full shutdown of the system and are stored in a fault memory. This can be interrogated in the garage-workshop with a diagnosis tester and deleted after the faults have been cleared.

ABS system variants
A variety of versions are available depending on the brake-circuit configuration, the vehicle's drivetrain configuration, functional requirements and cost considerations. The most common brake-circuit configuration is the diagonal split (X brake-circuit configuration), followed by the front-rear split (II brake-circuit configuration) (see Brake-circuit configuration). The HI and HH brake-circuit configurations (e.g., in the DaimlerChrysler Maybach) are specialized applications and are rarely used in combination with ABS. The ABS system variants are distinguished according to the number of control channels and wheel-speed sensors.

4-channel system with four sensors
These systems (Figure 10) allow individual control of the wheel brake pressure at each wheel by the four hydraulic channels, with the brake circuits split front/rear (for type-II brake-circuit configuration) or diagonally (for type-X brake-circuit configuration). Each wheel has its own wheel-speed sensor to monitor wheel speed.

Figure 9: Schematic diagram of an ABS/TCS hydraulic circuit for passenger cars with X brake-circuit configuration
1 Brake master cylinder, 2 Wheel-brake cylinder, 3 Hydraulic unit, 4 Intake valve, 5 Changeover valve, 6 Inlet valve, 7 Outlet valve, 8 Accumulator, 9 Return pump, 10 Pump motor.
R Right, L Left,
F Front, R Rear.



3-channel system with three sensors
Instead of the familiar arrangement with a separate speed sensor on each wheel, the rear wheels with this variant share a single sensor which is fitted in the differential. Due to the characteristics of the differential, it allows the measurement of wheel-speed differences with certain restrictions. Due to the select-low control characteristic for the rear wheels, i.e., parallel connection of the two rear-wheel brakes, a single hydraulic channel is sufficient for (parallel) control of the rear braking pressures.

Hydraulic 3-channel systems require a type-II brake-circuit configuration (front and rear split).

3-sensor systems can only be used on vehicles with rear-wheel drive, i.e., primarily small commercial vehicles and light trucks. The number of vehicles fitted with such systems is generally dropping off.

2-channel system with one or two sensors
Two-channel systems were produced because of the smaller number of components required and the resulting potential for cost savings. Their popularity was limited as their functionality does not match that of "full-fledged" systems. These systems are now hardly ever used in cars.

TCS engine-control intervention in passenger cars
On passenger cars with diesel engines, engine-control intervention is performed, depending on the variant, by means of electronic diesel control or the ETC (electronic throttle control) system (by reduced delivery).

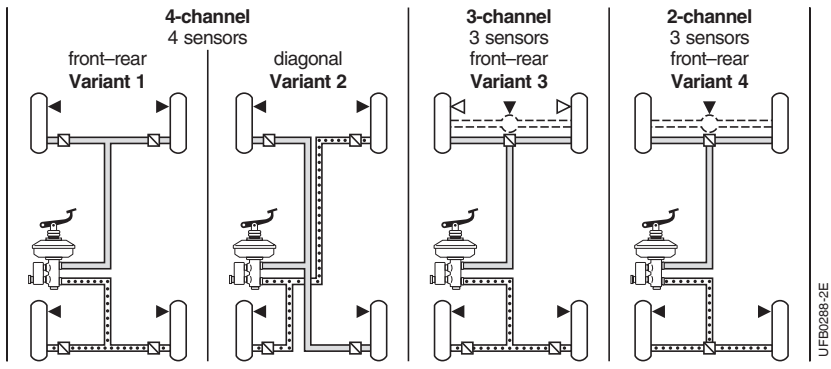
On passenger cars with gasoline engines, torque reduction is usually performed by means of a combination of several functions. In this way, it is possible to reduce the engine torque according to the requirements by specifically suppressing injection pulses, retarding the ignition timing or closing the throttle device (ETC).

The engine-management systems receive the TCS request via signal or CAN data lines from the TCS control system.

Use of ABS on motorcycles
It has been possible to reduce the size and weight of car ABS systems substantially in recent years. As a result, volume-production ABS systems are now a very attractive option for motorcycles. Consequently, this class of vehicle will also be able to benefit from the advantages of ABS as a safety system.

The car system is modified for use on motorcycles. Instead of the usual eight 2/2-way valves in the hydraulic unit for cars (with X-configuration brake circuits), motorcycles normally only require four valves. The control algorithm also differs fundamentally from that of a car ABS system.

Figure 10: ABS system variants
◹ Control channel, ◀ Sensor, ◁ Sensor (alternative to differential sensor).



Other system variants have arisen from the demand for combined brake systems (CBS), i.e., systems in which both the front and rear brakes can be operated either by a foot pedal or a hand-operated lever, possibly in combination with a separate means of actuating the front brakes. This type of special case requires a 3-channel hydraulic unit. However, the CBS variant is very model-specific in design.

ABS/TCS systems for commercial vehicles

In an ABS or ABS/TCS system for commercial vehicles, depending on the vehicle configuration and the number of axles, four or six wheel-speed sensors, up to three to six pressure-control valves, and – in the case of a TCS system – one TCS valve are used (Figure 11 and Figure 12).

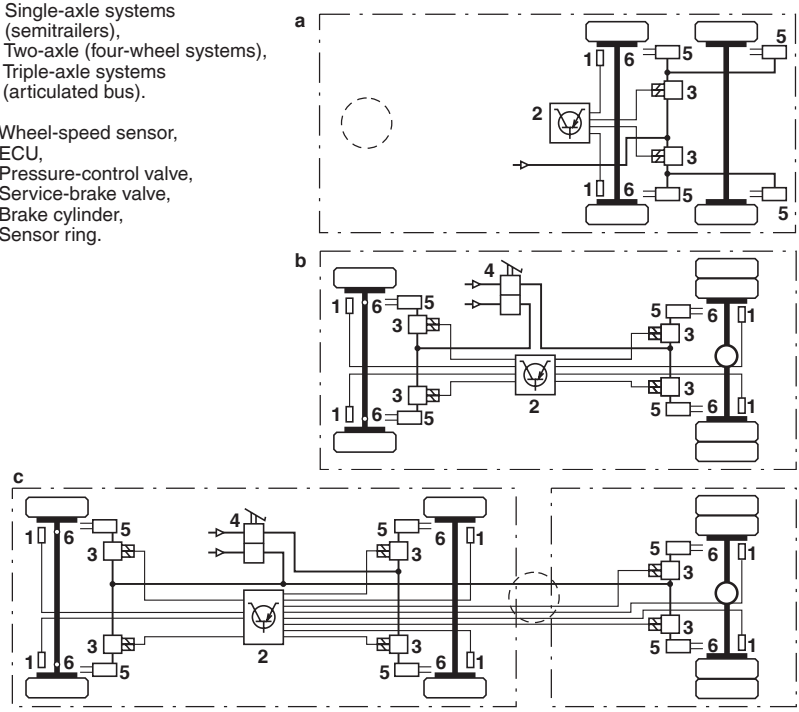
During a learning operation on initial commissioning, the control unit adjusts to the corresponding vehicle depending on the connected components. This involves detecting the number of axles, the ABS control method, and any additional functions that may be required, such as TCS.

If one axle is a lifting axle, it is automatically excluded from the ABS control process when lifted. When two axles are close together, often only one of them is fitted with wheel-speed sensors. The brake pressure of two wheels arranged

Figure 11: Examples of ABS systems for commercial vehicles

a) Single-axle systems (semitrailers),
b) Two-axle (four-wheel systems),
c) Triple-axle systems (articulated bus).

1 Wheel-speed sensor,
2 ECU,
3 Pressure-control valve,
4 Service-brake valve,
5 Brake cylinder,
6 Sensor ring.



one after the other is then controlled jointly by a single pressure-control valve. In multi-axle vehicles where the axles are further apart, such as articulated buses for example, three-axle control is used out of preference.

In light commercial vehicles with pneumatic-hydraulic converters, ABS intervenes in the pneumatic brake circuit via pressure-control valves and defines the hydraulic braking pressure.

When the vehicle is running on a low-friction-coefficient road surface, the operation of an additional retarding brake (exhaust brake or retarder) can lead to excessive slip at the driven wheels. This would impair vehicle stability. ABS therefore also monitors brake slip here and controls it to permissible levels by switching the retarder on and off.

In addition, there is an ABS system independent of the tractor vehicle in the trailer, which in turn consists of two or four wheel-speed sensors and an mechatronic pressure-control module with integrated electronics.

Components

Wheel-speed sensors

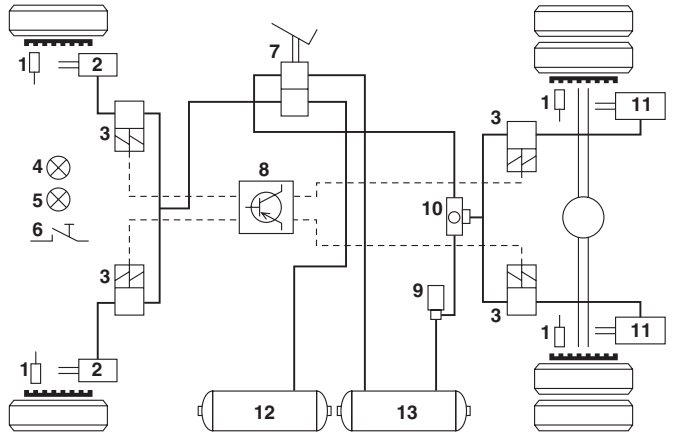
Virtually exclusively passive speed sensors in accordance with the inductive measuring principle have hitherto been used in the commercial-vehicle sector. Owing to the principles involved, these sensors can only sense speeds greater than 0 km/h, which means that – unlike active sensors, which can also detect a stationary sensor wheel – stationary detection is not possible.

Pressure-control valve

The pressure-control valves are located between the service-brake valve and the brake cylinders, and control the brake pressure of one or more wheels (Figure 13). As pilot-controlled valves, they consist of a combination of two solenoid valves and in each case a downstream pneumatic diaphragm valve, which are designed as outlet and holding valves (single-channel pressure-control valve). The electronics control the solenoid valves in the appropriate combination so that the required function is performed ("pressure holding" and "pressure drop"). If no pilot-valve actuation takes place, "pressure rise" is the result.

Figure 12: Traction control system for commercial vehicles

1 Wheel-speed sensor with sensor ring, 2 Brake cylinder, 3 ABS pressure-control valve, 4 ABS warning lamp, 5 TCS lamp, 6 TCS switch, 7 Service-brake valve, 8 ABS/TCS ECU, 9 TCS valve, 10 Shuttle valve, 11 Spring-type brake cylinder, 12 Compressed-air reservoir, circuit 1, 13 Compressed-air reservoir, circuit 2.



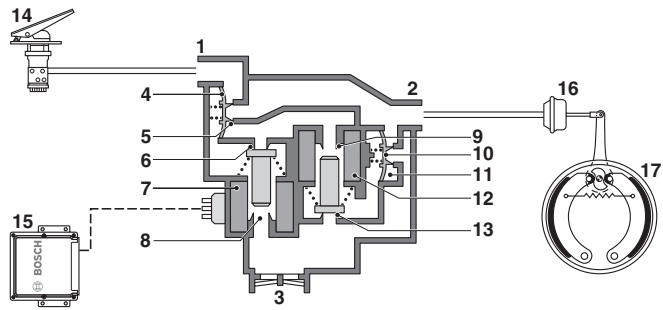
When braking normally (i.e., without ABS response = no incipient wheel locking), air flows through the pressure-control valves unhindered in both directions when pressure is applied to or vented from the brake cylinders. This ensures fault-free functioning of the service-brake system.

TCS valve
Usually a directly controlled TCS solenoid valve designed as a 2/2-way directional-control valve in combination with a pneumatic shuttle valve is provided for the driver-independent pressure rise for the TCS brake controller (Figure 12). In the event of a brake intervention the supply pressure is applied via the electrically actuated TCS solenoid valve and the shuttle valve to the ABS pressure-control valves, whereby the shuttle valve blocks the connection to the service-brake valve according to select-high. At the same time, on the non-active side, the holding magnet of the ABS pressure-control valve is actuated to prevent the pressure rise in the associated wheel-brake cylinder. The brake pressure can now be controlled according to the desired wheel speed with the ABS pressure-control valve on the active side.

TCS engine-control intervention in commercial vehicles
Engine-control intervention is performed – similarly to passenger cars with diesel engines – via electronic diesel control (e.g., by reduced delivery). The engine-management system receives the corresponding signal via a CAN data bus from the ABS/TCS ECU.

References
[1] ECE-R13: Uniform provisions concerning the approval of category M, N and O vehicles with regard to brakes.

Figure 13: Pressure-control valve
1 Connection for energy input, 2 Connection for energy output, 3 Venting, 4 Diaphragm, 5 Inlet, 6 Valve seat, 7 Solenoid valve for pressure-holding valve, 8 Valve seat, 9 Valve seat, 10 Diaphragm, 11 Outlet, 12 Solenoid valve for outlet valve, 13 Valve seat, 14 Service-brake valve, 15 ABS/TCS ECU, 16 Pressure-control valve, 17 Wheel brake.



Driving-dynamics control system

Function

Human error is the cause of a large portion of road accidents. Even under normal driving conditions, a driver and his vehicle can reach their physical operating limits on account of, for example, an unexpected bend in the road, a suddenly appearing obstacle or an unanticipated change in the condition of the road surface. Increased speed can also result in the driver not being able to control his vehicle safely, since the lateral-acceleration forces acting on the vehicle in such a situation reach levels which make excessive demands on him.

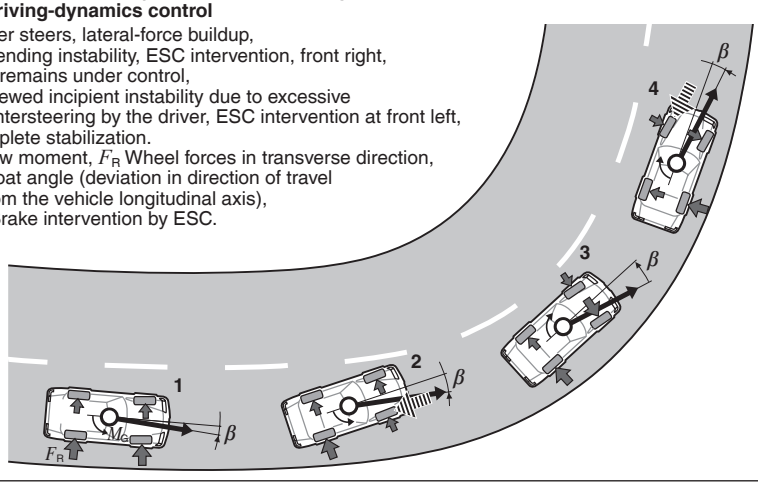
If the tires' coefficients of friction are exceeded, the vehicle will suddenly behave differently from what the driver with his driving experience expects. In such situations of operating limits, the driver is often no longer able to stabilize the vehicle himself; as a rule, he will intensify the instability through reactions arising from fear and panic. As a result, a significant discrepancy is built up between the longitudinal motion of the vehicle and the longitudinal axis of the vehicle (float angle β). Even by steering in the opposite direction, a normal driver will barely be able to restabilize his vehicle on his own at float angles in excess of 8°.

Driving-dynamics control is known by different names – the Electronic Stability Program (ESP®, brand name of Daimler AG) or the neutral designation Electronic Stability Control (ESC). The system makes a substantial contribution to neutralizing such limit situations by helping the driver to keep his/her vehicle under control within the framework of the physical limits. Sensors constantly record the behavior of both the driver and the vehicle. By comparing the actual state with a target state appropriate for the relevant situation, the system, in the event of significant discrepancies, makes interventions in the braking system and in the drivetrain to stabilize the motion of the vehicle (Figure 1).

The integrated functionality of the antilock braking system (ABS) prevents the wheels from locking when the brakes are applied, while the similarly integrated traction control system (TCS) inhibits excessive wheel spin during acceleration. ESC as an overall system, however, embraces capabilities extending far beyond those of either ABS or ABS and TCS combined. The system ensures that the vehi-

Figure 1: Transverse dynamics for a passenger car with driving-dynamics control

1 Driver steers, lateral-force buildup,
2 Impending instability, ESC intervention, front right,
3 Car remains under control,
4 Renewed incipient instability due to excessive countersteering by the driver, ESC intervention at front left, complete stabilization.
M_G Yaw moment, F_R Wheel forces in transverse direction,
β Float angle (deviation in direction of travel from the vehicle longitudinal axis),
Brake intervention by ESC.



cle does not swerve outwards with its rear end (oversteering) or does not push excessively outwards with its front end (understeering), but instead follows the driver's steering input as far as is physically possible.

ESC is based on ABS and TCS components. In this way, the individual wheels can be actively braked with high dynamic response. The engine torque and thus the traction-slip values at the wheels can be influenced by means of the engine-management system. The systems communicate via, for example, the CAN bus.

Table 1: Terms and quantities

a_y	Measured vehicle lateral acceleration
F_x	Tire force in longitudinal direction
F_y	Tire force in lateral direction (lateral force)
F_N	Tire force in normal direction (normal force)
L	Distance between front and rear axles
M_{BrNom}	Nominal braking torque
M_{DifNom}	Nominal differential torque
M_{EngNom}	Nominal engine torque
$M_{MWhlNom}$	Nominal sum torque
ΔM_{RedNom}	Setpoint change of engine-torque reduction
ΔM_Z	Stabilizing yaw moment
p_{Whl}	Wheel-cylinder pressure
p_{Adm}	Admission pressure
r	Radius of bend
v_{ch}	Characteristic vehicle speed
v_{Dif}	Wheel-speed differential of drive wheels (on one axle)
v_{DifNom}	Nominal wheel-speed differential of drive wheels (on one axle)
v_{MWhl}	Mean wheel speed of driven axle
$v_{MWhlNom}$	Setpoint value of mean wheel speed
v_{Whl}	Measured wheel speed
v_x	Vehicle linear velocity
v_y	Vehicle lateral velocity
α	Tire slip angle
β	Float angle
δ	Steering-wheel angle
λ	Tire slip
λ^i_{Nom}	Setpoint value of tire slip at wheel i
$\Delta\lambda_{DifTolNom}$	Setpoint change of permissible slip differential of driven axle(s)
$\Delta\lambda_{Nom}$	Setpoint slip change
μ	Coefficient of friction
$\dot{\psi}$	Yaw velocity
$\dot{\psi}_{Nom}$	Nominal yaw velocity

Requirements

Driving-dynamics control contributes to increasing driving safety. It improves vehicle behavior up to the physical operating limits. The vehicle's reaction remains foreseeable to the driver and can thus be better controlled even in critical driving situations.

At the vehicle's physical driving limits, vehicle and directional stability are enhanced in all operating states, such as full braking, partial braking, coasting, accelerating, overrunning and load changes, and also for example in the case of extreme steering maneuvers (fear and panic reactions). The risk of skidding is drastically reduced.

In a variety of different situations, further improvements are obtained in the exploitation of traction potential when ABS and TCS come into action, and when engine drag-torque control is active (automatic increase in engine speed to inhibit excessive engine braking torque). This leads to shorter braking distances and greater traction with enhanced stability and higher levels of steering response.

Incorrect system interventions could have an impact on safety. A comprehensive safety concept ensures that all faults that are not essentially avoidable are detected in time and the ESC system is shut down fully or partially depending on the type of fault.

Numerous studies (e.g. [1] and [2]) have demonstrated that ESC drastically reduces the number of accidents caused by skidding and the number of associated fatalities. The upshot of this is that ESC will become mandatory in vehicles in North America by September 2011. In the European Union (EU), all new passenger cars and light commercial vehicles have had to be fitted with a driving-dynamics control system since November 2011 (integral part of ECE-R 13-H [12]). Other new cars were subject to a transition period which must be complied with by the end of 2014. Other regions such as Japan and Australia for example will also be introducing or have already introduced such a control system.

Operating principle

Driving-dynamics control (ESC) is a system which uses a vehicle's brake system and drivetrain to deliberately influence the vehicle's longitudinal and lateral motion in critical situations. When the stability-control function assumes operation, it shifts the priorities that govern the brake system. The basic function of the wheel brakes – to decelerate and/or stop the vehicle – assumes secondary importance as ESC intervenes to keep the vehicle stable and on course. ESC can also accelerate the drive wheels by means of engine interventions to contribute to the vehicle's stability.

Both mechanisms act on the vehicle's intrinsic motion. During steady-state circular-course driving, there is a defined connection between the driver's steering input and the resulting vehicle lateral acceleration and thus the tire forces in the lateral direction (self-steering effect). The forces acting on a tire in the longitudinal and lateral directions are dependent on the tire slip. It follows from this that the vehicle's intrinsic motion can be influenced by the tire slip. The specific braking of individual wheels, e.g. of the rear wheel on the inside of the bend in the case of understeering or of the front wheel on the outside of the bend in the case of oversteering, helps the vehicle to remain on the course determined by the steering angle as precisely as possible.

Typical driving maneuver

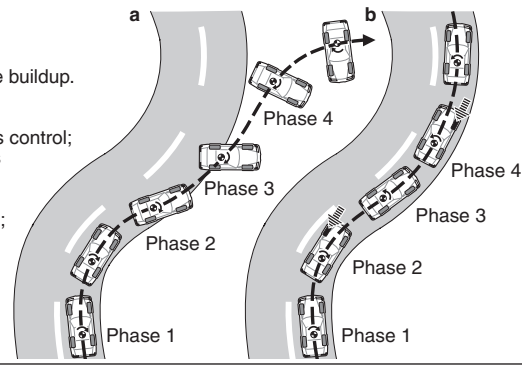
To compare how a vehicle handles at its operating limits with and without ESC, the following example is given. The driving maneuver reflects actual operating conditions, and is based on simulation programs designed using data from vehicle testing. The results have been confirmed in subsequent road tests.

Rapid steering and countersteering
Figure 2 demonstrates the handling response of a vehicle without ESC and of a vehicle with ESC negotiating a series of S-bends with rapid steering and countersteering inputs on a high-grip road surface (coefficient of friction $\mu = 1$), without the driver braking and at an initial speed of 144 km/h. Figure 3 shows the curves for dynamic-response parameters. Initially, as they approach the S-bend, the conditions for both vehicles, and their reactions, are identical. Then come the first steering inputs from the drivers (phase 1).

Vehicle without ESC
As can be seen, in the period following the initial, abrupt steering input the vehicle without ESC is already threatening to become unstable (Figure 2a, Phase 2). Whereas the steering input has quickly generated substantial lateral forces at the front wheels, there is a delay before the rear wheels start to generate similar forces. The vehicle reacts with a clockwise

Figure 2: Vehicle tracking during right-left cornering sequence
a) Vehicle without ESC,
b) Vehicle with ESC.
Brake intervention by ESC.

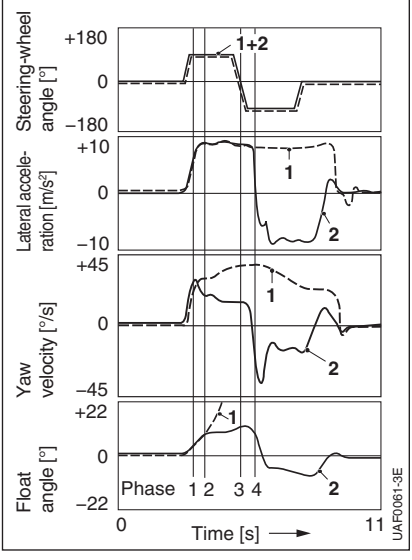
Phase 1: Driver steers, lateral-force buildup.
Phase 2: Incipient instability.
Phase 3: Countersteering:
vehicle without ESC loses control;
vehicle with ESC remains under control.
Phase 4: Vehicle without ESC is no longer under control;
Vehicle with ESC fully stabilized by ESC intervention at front right.



movement around its vertical axis (inward yaw). The vehicle barely responds to the driver's attempt to countersteer (second steering input, Phase 3), because it is no longer under control. The yaw velocity and the side-slip angle rise radically, and the vehicle breaks into a skid (phase 4).

Vehicle with ESC
The vehicle with ESC is stabilized after the initial steering input by active braking of the front left wheel to counter the threat of instability (Figure 2b, Phase 2): This occurs without any intervention on the driver's part. This action limits the inward yaw with the result that the yaw velocity is reduced and the float angle is not subject to an uncontrolled increase. Following the change of steering direction, first the yaw moment and then the yaw velocity reverse their directions (between Phases 3 and 4). In Phase 4, a second brief brake application – this time at the right front wheel – restores complete stability. The vehicle remains on the course defined by the steering-wheel angle.

Figure 3: Dynamic-response curves during a right-left cornering sequence
1 Vehicle without ESC, 2 Vehicle with ESC. Phases 1-4 see Figure 2.



Structure of the overall system

Objective of driving-dynamics control
The control of the handling characteristics at the vehicle's physical driving limits is intended to keep the vehicle's three degrees of freedom in the plane of the road – linear velocity v_x, lateral velocity v_y and yaw velocity $\dot{\psi}$ about the vertical axis – within the controllable limits. Assuming appropriate operator inputs, driver demand is translated into dynamic vehicular response that is adapted to the characteristics of the road in an optimization process designed to ensure maximum safety.

System and control structure
The ESC system comprises the vehicle as a controlled system, the sensors for determining the controller input variables, the actuators for influencing the braking, motive and lateral forces, as well as the hierarchically structured controller, comprising a higher-level transverse-dynamics controller and lower-level wheel controllers (Figure 4). The higher-level controller determines the setpoint values for the lower-level controller in the form of moments or slip, or their changes. Internal system variables that are not directly measured, such as the float angle β for example, are determined in the driving-condition estimation ("observer").

In order to determine the nominal behavior, the signals defining driver command are evaluated. These comprise the signals from the steering-wheel-angle sensor (driver's steering input), the brake-pressure sensor (desired deceleration input, obtained from the brake pressure measured in the hydraulic unit) and the accelerator-pedal position (desired drive torque). The calculation of the nominal behavior also takes into account the utilized friction-coefficient potential and the vehicle speed. These are calculated in the observer from the signals sent by the wheel-speed sensors, the lateral-acceleration sensor, the yaw sensor, and the brake-pressure sensor. Depending on the control deviation, the yaw moment, which is necessary to make the actual-state variables approach the desired-state variables, is then calculated.

In order to generate the required yaw moment, it is necessary for the changes in desired braking torque and slip at the wheels to be determined by the transverse-dynamics controller. These are then set by means of the lower-level brake-slip and traction controllers together with the brake-hydraulics actuator and the engine-management actuator.

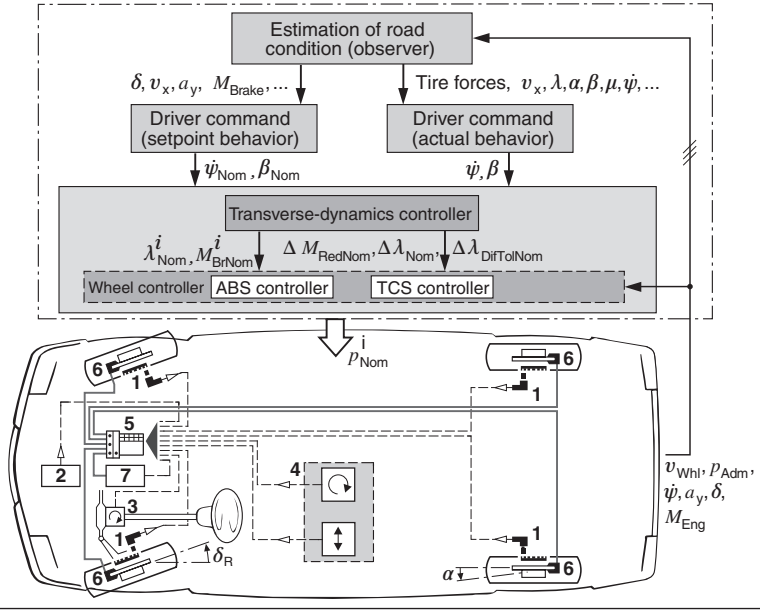
Driving-condition estimation

To determine the stabilization interventions, not only knowledge of the signals from the sensors for wheel speeds v_{Whl}, admission pressure p_{Adm}, yaw rate (yaw velocity) $\dot{\psi}$, lateral acceleration a_y, steering-wheel angle δ and engine torque is important, but also knowledge of a series of further internal system variables which can be measured indirectly with appropriate effort. These include, for example, the tire forces in the longitudinal, lateral and normal directions (F_x, F_y and F_N), the vehicle linear velocity v_x, the tire slip values λ_i, the slip angle α on one axle, the float angle β, the vehicle lateral velocity v_y, and the coefficient of friction μ. They are estimated on a model-supported basis from the sensor signals in the observer.

The vehicle linear velocity v_x is of crucial importance to all the wheel-slip-based controllers and must therefore be calculated with great accuracy. This is done on the basis of a vehicle model using the measured wheel speeds. Numerous influences must be taken into account here. The vehicle speed v_x differs for example already in normal driving situations on account of brake or drive slip from the wheel speeds v_{Whl}. For all-wheel-drive vehicles, special linking of the wheels must be taken into account in particular. During cornering, the

Figure 4: ESC overall control system

1 Wheel-speed sensors,
2 Brake-pressure sensor (integrated in hydraulic unit),
3 Steering-wheel-angle sensor,
4 Yaw sensor (yaw-rate sensor) with integrated lateral-acceleration sensor,
5 ESC hydraulic unit (hydraulic modulator) with mounted ECU,
6 Wheel brakes, 7 Engine ECU.



wheels on the inside of the bend follow a different course from the wheels on the outside of the bend, and consequently rotate at a different speed.

Vehicle handling changes during normal use in response to varying load, altered tractive resistance (e.g. road gradient or surface, wind) or wear (e.g. of the brake pads).

Under all these boundary conditions, the vehicle linear velocity must be estimated with a deviation of a few % in order to ensure the enabling and intensity of stabilization interventions to the necessary extent.

Basic transverse-dynamics controller

The function of the transverse-dynamics controller is to calculate the actual behavior of the vehicle from, for example, the yaw-velocity signal and the float angle estimated in the observer, and to bring the driving behavior in the driving-dynamic limit range into line with behavior in the normal range as closely as possible (nominal or setpoint behavior).

The connection that exists during steady-state circular-course driving between the yaw velocity and the steering-wheel angle δ, the vehicle linear velocity v_x and characteristic vehicle variables are used to determine the nominal behavior. The single-track vehicle model (see e.g. [3]) is used to produce

$$\dot{\psi} = \frac{v_x}{l} \; \delta \; \frac{1}{1+\left(\frac{v_x}{v_{ch}}\right)^2}$$

as the basis for calculating the vehicle nominal motion. In this formula, l denotes the distance between the front and rear axles. Geometric and physical parameters of the vehicle model are summarized in the "characteristic vehicle speed" v_{ch}.

The variable $\dot{\psi}$ is then limited according to the current friction-coefficient conditions and to the special properties of the vehicle dynamics and the driving situation (e.g. braking or acceleration by the driver) and to the particular conditions such as a sloping road surface of different friction coefficients under the vehicle (μ split). The driver command is thus known as the nominal yaw velocity $\dot{\psi}_{Nom}$.

The transverse-dynamics controller compares the measured yaw velocity with the associated setpoint value and in the event of significant deviations calculates the yaw moment that is required to match the actual state variable to its setpoint state. At a higher level, the float angle β is monitored and, as the values rise, increasingly taken into consideration in the calculation of the stabilizing yaw moment ΔM_z. This controller output variable is applied by means of braking-torque and slip inputs to the individual wheels which must be adjusted by the lower-level wheel controllers.

Stabilization interventions are performed at the wheels, the braking of which generates a yaw moment in the required direction of rotation and at which the limit of the transmittable forces has not yet been reached. For an oversteering vehicle, the physical limit is first exceeded on the rear axle. Stabilization interventions are therefore performed via the front axle. For an understeering vehicle, the situation is reversed (see e.g. [6]).

The nominal slip values requested by the transverse-dynamics controller λ^i_{Nom} at individual wheels are set with the aid of lower-level wheel controllers (see Figure 4). A distinction is made between the following three application cases.

Wheel control in the coasting case

In order to exert as accurately as possible the yaw moments required to stabilize the vehicle, the wheel forces must be altered under defined conditions by controlling the wheel slip. The nominal slip requested by the transverse-dynamics controller at a wheel is adjusted in the unbraked case by the lower-level brake-slip controller by way of an active pressure build-up. The current slip at the wheel must be known as precisely as possible for this purpose. This is calculated from the measured wheel-speed signal and the vehicle linear velocity determined in the observer v_x. The nominal braking torque at the wheel is formed from the deviation of the actual wheel slip from its setpoint value using a PID control law.

It is not only in the event of an active pressure build-up by transverse-dynamics control that a wheel can be subject to

brake slip. Following downshifts and when the accelerator is suddenly released, inertia in the engine's moving parts exerts a degree of braking force at the drive wheels. Once this force and the corresponding reactive torque rise beyond a certain level, the tires will lose their ability to transfer the resulting loads to the road and will tend to lock (e.g. because the road is suddenly slippery). The brake slip for the driven wheels can be limited in the coasting case by engine drag-torque control. This acts like "gentle acceleration" by the driver.

Wheel control in the braked case
In the braked case, different actions overlap at individual wheels, depending on the driving situation:
– driver input via the brake pedal and the steering wheel,
– effect of the ABS controller, which prevents individual wheels from locking,
– interventions of the transverse-dynamics controller which ensures vehicle stability by specifically braking individual wheels if necessary.

These three requirements must be coordinated in such a way that the driver's deceleration and steering inputs are implemented as much as possible. If wheel control is performed primarily with the objective of maximum vehicle deceleration, it can be performed on the basis of wheel acceleration which can be robustly determined with minimal sensor information (instability control). In order to specifically adjust the longitudinal and lateral tire forces to stabilize the vehicle, the principle of slip control [4] must be applied because it also permits wheel control in the unstable range of the friction-coefficient/slip characteristic. From the available sensor information, however, it must be possible to determine the absolute wheel slip to a few %, depending on the vehicle speed.

The function of the ABS controller is to ensure vehicle stability and steerability in all road conditions and in so doing exploit as much as possible the friction between wheels and road. It does this also in its capacity as the lower-level controller to the transverse-dynamics controller by modulating the brake pressure at the wheel in such a way that the maximum possible longitudinal force can be exerted while maintaining sufficient lateral stability. However, more variables are measured in ESC than in a pure ABS configuration, which only contains the wheel-speed sensors. Thus, individual vehicle-motion information, such as for example yaw rate or lateral acceleration, is available through direct measurement of greater accuracy than is the case with model-supported estimation of the basis of few measured values.

In certain situations, it is possible to increase performance by adapting ABS control by means of inputs from the transverse-dynamics controller. When a vehicle decelerates on unequal road surfaces (μ split), very different braking forces occur at the wheels on the left and right sides of the vehicle. This generates a yaw moment about the vehicle vertical axis, to which the driver must react by countersteering in order to stabilize the vehicle. How quickly this yaw moment is built up – and how fast the driver must consequently react – depends on the vehicle's moment of inertia about the vertical axis. ABS features yaw-moment build-up delay to hold back the pressure rise at the front wheel with the higher coefficient of friction at the road surface ("High" wheel). This ABS facility can also use information from the higher-level transverse-dynamics controller (on the driver's reaction and the vehicle behavior) and thereby react even better to the actual vehicle motion.

If, when braking in a bend, the vehicle starts to turn under certain conditions, the oversteer tendency can be counteracted by means of electronic braking-force distribution through pressure reduction in individual wheels. If this is not sufficient on its own, the transverse-dynamics controller helps by actively building up pressure at the front wheel on the outside of the bend (reduction of lateral force). If, on the other hand, the vehicle understeers, the braking torque is increased at the rear on the inside (provided the wheel is not subject to ABS control) and decreased slightly at the front on the outside.

If the vehicle starts to oversteer during a fully or partially braked lane change, the pressure at the rear wheel on the inside of the bend is specifically decreased (increase in lateral force) and the pressure at the front wheel on the outside of the bend increased (decrease in lateral force). If the vehicle understeers while braking in the bend, the braking torque is increased at the rear on the inside (provided the wheel is not yet in the ABS control range) and decreased slightly at the front on the outside.

Wheel control in the drive case

The lower-level traction controller (TCS) is activated as soon as the drive wheels start to spin in the drive case. The measured wheel speed and the respective drive slip can be influenced by changing the torque balance at each drive wheel. The TCS controller limits the drive torque at each drive wheel to the drive torque that can be transferred there to the road. In this way, the driver command is implemented after acceleration as well as is physically possible and, at the same time, fundamental directional stability is ensured, since the lateral forces at the wheel are not too greatly reduced.

In a vehicle with a powered axle, the mean wheel speed of the driven axle

$$v_{\mathrm{MWhl}} = \frac{1}{2}\left(v^{\mathrm{L}}_{\mathrm{Whl}} + v^{\mathrm{R}}_{\mathrm{Whl}}\right)$$

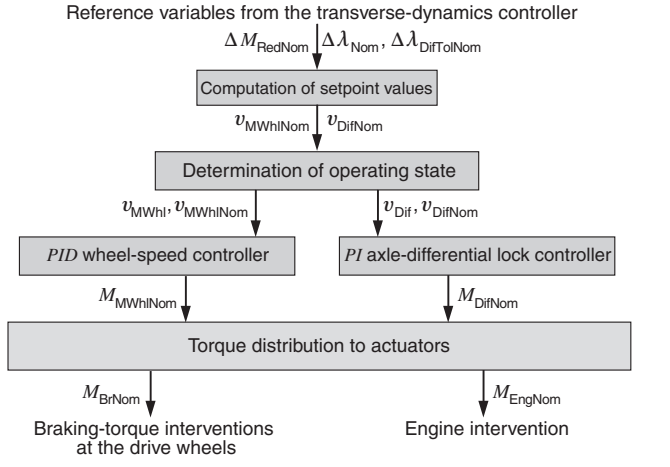
and the wheel-speed differential

$$v_{\mathrm{Dif}} = v^{\mathrm{L}}_{\mathrm{Whl}} - v^{\mathrm{R}}_{\mathrm{Whl}}$$

between the measured wheel speed of the left wheel $v^{\mathrm{L}}_{\mathrm{Whl}}$ and the right wheel $v^{\mathrm{R}}_{\mathrm{Whl}}$ are used as controlled variables.

The structure of the entire TCS controller is depicted in Figure 5. The reference variables of the transverse-dynamics controller are included in the setpoint-value calculation for the mean wheel speed and the wheel-speed differential, as well as the setpoint slip values and the coasting wheel speeds. In the calculation of the setpoint values v_{DifNom} (setpoint wheel-speed differential of the drive wheels on one axle) and v_{MWhlNom} (setpoint value of the mean wheel speed), the inputs for changing the setpoint slip $\Delta\lambda_{\mathrm{Nom}}$ and the permissible slip differential $\Delta\lambda_{\mathrm{DifTolNom}}$ of the drive axle(s) act in the form of an offset on the basic values calculated in the TCS. In addition, an understeer or oversteer tendency identified by the transverse-dynamics controller directly influences, through the setpoint change of the en-

Figure 5: Structure of TCS controller
Quantities, see Table 1.



gine-torque reduction ΔM_{RedNom}, the determination of the maximum permissible drive torque.

The dynamic response of the drivetrain depends on the highly differing operating states. It is therefore necessary to determine the current operating status (e.g. selected gear, clutch actuation) in order to be able to adapt the controller parameters to the controlled system's dynamic response and to nonlinearities.

Because the mean wheel speed is affected by variable inertial forces originating from the drivetrain as a whole (engine, transmission, drive wheels, and the propshaft itself), a relatively large time constant is employed to describe its correspondingly leisurely rate of dynamic response. The mean wheel speed is controlled by means of a nonlinear PID controller, whereby in particular the gain of the I-component (dependent on the operating status) can vary over a wide range. In the stationary case, the I-component is a measure for the torque which can be transferred to the road surface. The output variables of this controller is the setpoint sum torque $M_{MWhlNom}$.

In contrast, the time constant for the wheel-speed differential is relatively small, reflecting the fact that the wheels' own inertial forces are virtually the sole determining factor for their dynamic response. Furthermore, in contrast to the mean wheel speed, it is influenced only indirectly by the engine. The wheel-speed differential v_{Dif} is controlled by a nonlinear PI controller. Because brake interventions at a drive wheel initially only become noticeable through the torque balance of this wheel, they change the distribution ratio of axle differential and thereby emulate a differential lock. The controller parameters of this axle-differential-lock controller are only dependent on the engaged gear and engine influences to a minimal extent. If the differential speed on the driven axle deviates more than currently permissible (dead zone) from its setpoint value v_{DifNom}, calculation of a setpoint differential torque M_{DifNom} starts. The dead zone is widened if TCS brake interventions are to be avoided, for example when cornering at the operating limits.

Setpoint sum and setpoint differential torques are the basis for distributing the positioning forces to the actuators. The setpoint differential torque M_{DifNom} is set by the braking-torque difference between the left and right drive wheels by means of corresponding valve actuation in the hydraulic unit (asymmetrical brake intervention). The setpoint sum torque $M_{MWhlNom}$ is adjusted by both the engine interventions and a symmetrical brake intervention.

With a gasoline engine, adjustments undertaken through the throttle valve are relatively slow to take effect (lag and the engine's transition response). Retarding the ignition timing and, as a further option, selective suppression of injection pulses are employed for rapid engine-based intervention. In diesel-engine vehicles, the electronic diesel control system (EDC) reduces the engine torque by modifying the quantity of fuel injected. Symmetrical brake intervention can be applied for brief transitional support of engine-torque reduction.

Traction plays a special role in off-road applications. Normally, in vehicles with off-road requirements, traction control is automatically adapted by way of special situation identification in order to achieve the best levels of performance and robustness possible. Other vehicle manufacturers give the driver the opportunity to choose different adjustments from deactivation of engine-torque limitation through to adaptations tailored to special road conditions (e.g. ice, snow, grass, sand, slush, and rocky ground).

Supplementary transverse-dynamics functions

The basic ESC functions described above can also feature supplementary driving-dynamics functions for special vehicle categories, such as sport utility vehicles (SUVs) and small vans for example, and for special vehicle-stabilization requirements.

Enhanced understeering control

It is possible even in normal driving conditions for the vehicle to fail to comply adequately with the driver's steering input

(it understeers) if, for example, the road surface in a bend is suddenly wet or contaminated. ESC can therefore increase the yaw rate by exerting an additional yaw moment. This enables the vehicle to negotiate a bend at the maximum speed physically possible. The expected frequency of interventions and the comfort requirements of the vehicle manufacturer differ from vehicle type to vehicle type, and there are accordingly different expansion stages for executing such brake interventions which influence the vehicle's understeering behavior.

If the driver requests a smaller radius of bend than is physically possible, then only the reduction of the vehicle speed remains. This can be read from the connection applicable during steady-state cornering between the radius of bend r, the vehicle linear velocity v_x and the yaw rate $\dot{\psi}$:

$$r = \frac{v_x}{\dot{\psi}} .$$

In order to ensure a desired track course, the vehicle is then – without a yaw moment being applied – braked as far as necessary by specific braking of all the wheels (Enhanced Understeering Control, EUC).

Rollover prevention

In particular, light commercial vehicles and other vehicles with a high center of gravity, such as sport utility vehicles (SUVs), can overturn when high lateral forces are generated by a spontaneous steering reaction by the driver in the course of an evasive maneuver on a dry road for example (highly dynamic driving situations) or when the lateral acceleration of a vehicle slowly increases into the critical range as it negotiates a freeway exit with a decreasing radius of bend at excessive speed (quasi-stationary driving situations).

There are special functions (Rollover Mitigation Functions, RMF) which identify these critical driving situations by using the normal ESC sensors and stabilize the vehicle by intervening in brake and engine control. In order to ensure intervention on time, in addition to the driver's steering input and the measured reaction of the vehicle (yaw rate and lateral acceleration), a predictive process is used to estimate the vehicle's behavior in the near future. The two wheels on the outside of the bend, in particular, are braked if an imminent danger of overturning is identified. This action reduces the lateral forces on the wheels and thereby reduces the critical lateral acceleration. Particularly in the event of highly dynamic evasive maneuvers, wheel control must be effected with such high levels of sensitivity that, in spite of the wildly fluctuating vertical forces F_N, vehicle steerability is not diminished by the tendency of individual wheels to lock. The reduction of wheel speed by individual wheel braking also ensures that the driver is able to keep the vehicle in lane. In quasi-stationary driving situations, punctual reduction of the engine torque also prevents the driver from provoking a critical situation.

The moment of intervention and the intensity of the stabilizing interventions must be adapted as accurately as possible to the current vehicle behavior. This behavior can change significantly with the load, for example in the case of vans or sport utility vehicles fitted with roof racks. Such vehicles therefore make use of additional estimation algorithms which calculate the vehicle mass and the change in the center of mass caused by load distribution, if this is required to adapt the ESC functions (Load Adaptive Control, LAC).

Trailer sway mitigation

Depending on the vehicle speed, combinations of towing vehicle and trailer are prone to swaying about their vertical axis. If the vehicle is traveling at a slower speed than the "critical speed" (normally between 90 km/h and 130 km/h), these swaying motions are adequately damped and are quick to die down. If, however, the combination is traveling at a higher speed, small steering movements, crosswinds or driving over a pothole can suddenly induce such swaying motions, which then quickly intensify and can ultimately cause an accident due to the combination jackknifing.

Clear periodic oversteering triggers normal ESC stabilization interventions, but these normally arrive too late and on their own are not sufficient to stabilize the combination. The Trailer Sway Mitigation (TSM) function identifies swaying motions in good time on the basis of the customary ESC sensors; it does this by model-based analysis of the towing vehicle's yaw rate while taking into account the driver's steering movements. When these swaying motions reach a critical level, the combination is automatically braked in order to reduce the speed to such an extent that not even the smallest subsequent excitation will cause an immediate critical oscillation again. In order to damp the oscillation as effectively as possible in a critical situation, in addition to symmetrical deceleration through all the towing vehicle's wheels, individual wheel interventions are performed which swiftly damp the swaying motion of the combination. Limitation of the engine torque prevents dangerous acceleration by the driver during the stabilization process.

Activation of further driving-dynamics actuators

In addition to utilization of the hydraulic wheel brakes, other actuators are provided by means of which the driving-dynamic properties of a vehicle can be specifically influenced. When active steering and chassis systems are linked with ESC to form the composite system known as Vehicle Dynamics Management (VDM), they can, in their entirety, support the driver even better and thereby improve safety and driving dynamics even further.

While the combination of the steering or roll-stabilization system with the brake system has been introduced in the last few years [5], systems for activating differential locks in the drivetrain have been established on the market for some time now. The large number of such systems means that linking with ESC is possible in many cases. The supplementary actuator can basically be activated either directly from the extended ESC function (cooperation approach) or via a separate ECU which exchanges information with the ESC ECU (coexistence approach).

Figure 6: Drive concept of an all-wheel-drive vehicle with ESC

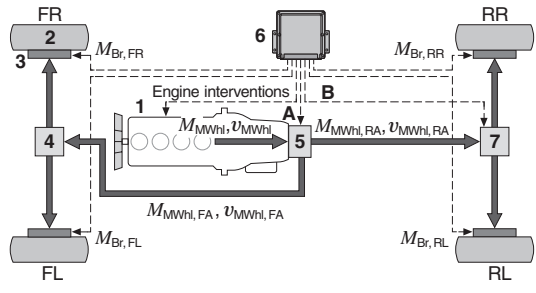
1 Engine with transmission,
2 Wheel,
3 Wheel brake,
4 Axle differential,
5 Central differential,
6 ECU with enhanced ESC function,
7 Axle differential.

Engine, transmission, gear ratios of differentials and losses are combined into one unit.

A Lock interventions with active central differential,
B Torque-vectoring interventions.

v Wheel speed,
v_{MWhl} Mean wheel speed,
M_{MWhl} Driving sum torque,
M_{Br} Braking torque,
FA Front axle
RA Rear axle

R Right, L Left,
F Front, R Rear,
FA Front axle,
RA Rear axle.



In all-wheel-drive vehicles, the drive torque is distributed via a central element to both powered axles (Figure 6). When the engine acts first and foremost on one axle and the second axle is linked via the central element, this is known as a hang-on system. If this central element is an open differential (without a locking action), drive torque is limited when one axle demonstrates increased slip. In the most unfavorable case, propulsion cannot be achieved if a wheel spins. In combination with ESC, symmetrical brake interventions by the all-wheel TCS controller can limit the differential speed between the axles and thereby achieve a longitudinal locking action.

The traction control of ESC can also be matched to the special operating concept of other types of central elements such as Torsen and viscous couplings. Basically, all the controllable drivetrain actuators must demonstrate a defined locking moment and dynamic response when opening and closing in order to specifically adapt the vehicle's self-steering properties with them.

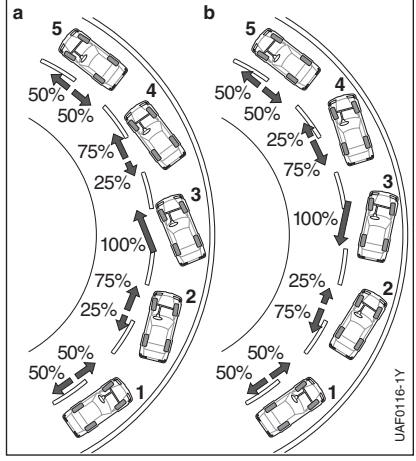
If the drivetrain of a vehicle can be manually switched over between different configurations, ESC can be automatically adjusted to the operating mode selected by the driver. Because ESC is based on individual wheel control, cooperation with mechanical differential locks for specific off-road conditions is only possible if the differential lock can be automatically opened during interventions by the transverse-dynamics controller. The system must otherwise be switched to an ABS fallback level when the lock is engaged, because driving-dynamics interventions at one wheel would also affect other wheels if the axles were rigidly linked.

In addition to simple links between the two axles, there are controllable central locks in which an electric or hydraulic actuator activates a coupling and thereby adapts the locking moment (Figure 6, A). In this way, it is possible with the ESC driving-dynamics information (e.g. wheel speeds, vehicle speed, yaw rate, lateral acceleration, and engine torque) and by also taking into account actuator-specific variables (such as e.g. the mechanical load on the component) to optimally adapt the linking of the two axles to the current driving situation (Dynamic Coupling Torque at Center, DCT-C).

The example in Figure 7 shows how variable drive-torque distribution influences the vehicle behavior. If, in the event of a risk of oversteering in a bend, provisionally more drive torque can be shifted to the front axle, it is necessary only much later to avoid instability, to lower the engine torque or even stabilize the vehicle with brake interventions (the maximum possible drive-torque shift is shown). If a vehicle tends to understeer, this tendency can be lessened by shifting drive torque to the rear axle. In both cases, vehicle behavior with improved response and better stability is achieved. The limits within which shifting the drive torque is actually possible are dependent on the concrete drivetrain configuration.

Figure 7: Influence of drive-torque distribution on vehicle behavior

a) Oversteer: Stability limit is first exceeded on the rear axle,
b) Understeer: Stability limit is first exceeded on the front axle.
1 Standard distribution during stable driving,
2 Incipient instability, drive torque is shifted to the axle which still has stability potential,
3 Maximum shift of drive torque,
4 Withdrawal of shift,
5 Standard distribution is re-established after instability has been reduced.



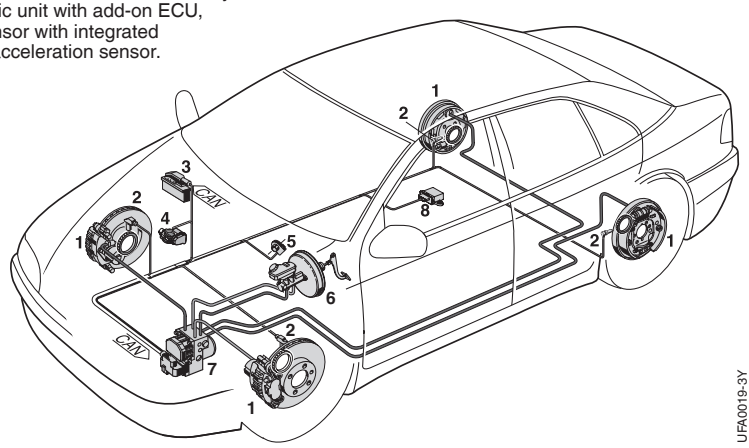
A controllable element on one axle can be activated by ESC along similar lines to the flexible linking of the two axles described. In terms of basic operation, the Dynamic Wheel Torque Distribution (DWT) function barely differs from the axle-differential lock effected by TCS via the hydraulic wheel brakes. However, such a supplementary actuator also actively distributes the drive torque between the wheels on one axle in normal situations. This is done with minimal losses and which much greater sensitivity and comfort than can be achieved by traction control in combination with braking-torque control and engine-torque reduction while taking into account the wear of the ESC hydraulic unit.

System components

The hydraulic unit, the ECU directly connected to it (add-on ECU) and the speed sensors are suitable for the rugged ambient conditions that are encountered in the engine compartment or the wheel arches. The yaw sensor and the lateral-acceleration sensor are either integrated into the ECU or like the steering-angle sensor installed in the passenger cell. Figure 8 shows by way of example the installation locations of the components in the vehicle with the electrical and mechanical connections.

Electronic control unit (ECU)
The ECU of printed-circuit board design comprises, as well as a dual-core computer, all the drivers and semiconductor relays for valve and pump activation, as well as interface circuits for sensor-signal conditioning and corresponding switch inputs for supplementary signals (e.g. brake-light switch). There are also interfaces (CAN, FlexRay) for communicating with other systems, such as engine and transmission management for example.

Figure 8: ESC components
1 Wheel brakes,
2 Wheel-speed sensors,
3 Engine ECU,
4 Electronically activated throttle valve,
5 Steering-angle sensor,
6 Brake booster with brake master cylinder,
7 Hydraulic unit with add-on ECU,
8 Yaw sensor with integrated lateral-acceleration sensor.



Hydraulic unit

The hydraulic unit (also called the hydraulic modulator), as in ABS or ABS/TCS systems, forms the hydraulic connection between the brake master cylinder and the wheel-brake cylinders. It converts the control commands of the ECU and uses solenoid valves to control the pressures in the wheel brakes. The hydraulic circuit is completed by bores in an aluminum block. This block is also used to accommodate the necessary hydraulic function elements (solenoid valves, plunger pumps and reservoir chambers).

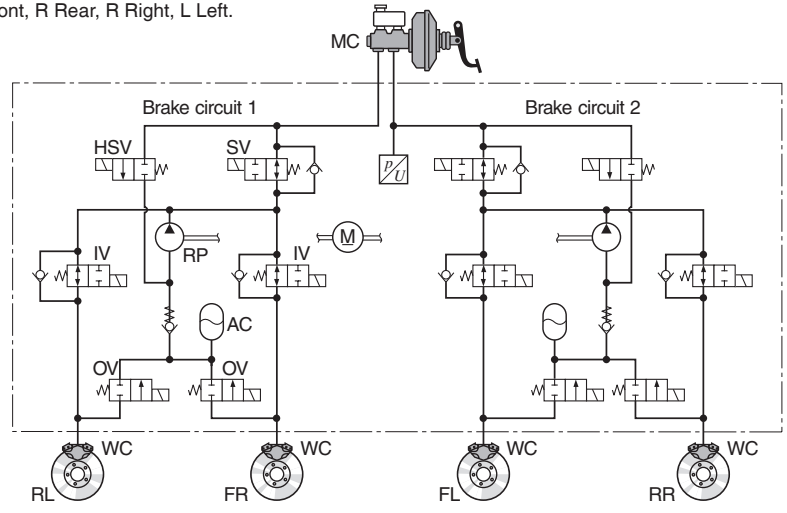
ESC systems require twelve valves irrespective of the brake-circuit configuration (Figure 9). In addition, a pressure sensor is usually integrated which measures the driver's deceleration command by way of the brake pressure in the brake master cylinder. This increases the performance of vehicle stabilization in partially active maneuvers. Pressure is modulated during ABS control (passive control) using ESC hydraulics in the same way as described for the ABS system.

But because ESC systems must also actively build up pressure (active control) or increase a brake pressure input by the driver (partially active control), the return pump used for ABS is replaced by a self-priming pump for each circuit. The wheel-brake cylinders and the brake master cylinder are connected via a changeover valve open at zero current and a high-pressure switching valve.

An additional non-return valve with a specific closing pressure prevents the pump from drawing unwanted brake fluid from the wheels. These pumps are driven by a DC motor based on demand. The motor drives an eccentric bearing located on the shaft of the motor.

Three examples of pressure modulation are shown in Figure 10. In order to build up pressure independently of the driver (Figure 10c), the switchover valves are closed and the high-pressure switching valves opened. The self-priming pump now pumps brake fluid to the relevant wheel or wheels in order to build up pressure. The inlet valves of the other wheels

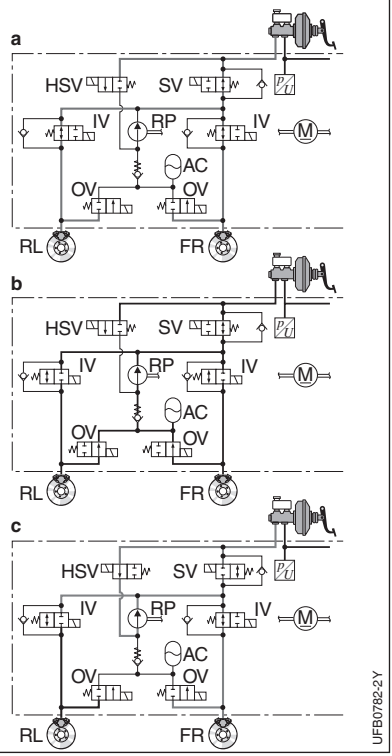
Figure 9: Hydraulic circuit diagram of an ESC hydraulic unit (X-brake-circuit configuration)
MC Brake master cylinder, WC Wheel-brake cylinder,
IV Inlet valve, OV Outlet valve, SV Switchover valve,
HSV High-pressure switching valve,
RP Return pump, M Pump motor,
AC Low-pressure reservoir,
F Front, R Rear, R Right, L Left.



remain closed. To reduce the pressure, the outlet valves are opened and the high-pressure switching valves and switchover valves return to their original position (Figure 10b). The brake fluid escapes from the wheels into the low-pressure reservoirs, which are run empty by the pumps. Demand-based control of the pump motor reduces noise emission during pressure generation and regulation.

Figure 10: Pressure modulation in the ESC hydraulic unit

a) Pressure build-up when braking,
b) Pressure reduction with ABS control,
c) Pressure build-up via self-priming pump due to TCS or ESC intervention.
IV Inlet valve,
OV Outlet valve,
SV Switchover valve,
HSV High-pressure switching valve,
RP Return pump, M Pump motor,
AC Low-pressure reservoir,
F Front, R Rear, R Right, L Left.



For partially active control (Figure 10a), the high-pressure switching valve must be able to open the suction path of the pump against higher differential pressures (> 0.1 MPa). The first stage of the valve is opened via the magnetic force of the energized coil; the second stage via the hydraulic area difference. If the ESC controller detects an unstable vehicle state, the switchover valves (open at zero current) are closed and the high-pressure switching valve (closed at zero current) is opened. The two pumps then generate additional pressure in order to stabilize the vehicle. When the intervention is finished, the outlet valve is opened and the pressure in the controlled wheel discharged to the reservoir. As soon as the driver releases the brake pedal, the fluid is pumped from the reservoir back to the brake-fluid reservoir.

Monitoring system

A comprehensive safety-monitoring system is of fundamental importance for reliable ESC functioning. The system used encompasses the complete system together with all components and all their functional interactions. The safety system is based on safety methods such as, for example, FMEA (Failure Mode and Effects Analysis), FTA (Fault Tree Analysis) and error-simulation studies. From these, measures are derived for avoiding errors which could have safety-related consequences. Extensive monitoring programs guarantee the reliable and punctual detection of all sensor errors which cannot be prevented completely. These programs are based on the well-proven safety software from the ABS and TCS systems which monitor all the components connected to the ECU together with their electrical connections, signals, and functions. The safety software was further improved by utilizing the possibilities offered by the additional sensors, and by adapting them to the special ESC components and functions.

The sensors are monitored at a number of stages. In the first stage, the sensors are continuously monitored during vehicle operation for line break, signal implausibility (“out-of-range” check), detection of interference, and physical plausibility. In a

second stage, the most important sensors are tested individually. The yaw-rate sensor is tested by intentionally detuning the sensor element and then evaluating the signal response. Even the acceleration sensor has internal background monitoring. When activated, the pressure-sensor signal must show a predefined characteristic, and the offset and amplification are compensated for internally. The steering-angle sensor has its own integrated monitoring functions and directly delivers a message to the ECU in the event of error. In addition, the digital signal transmission to the ECU is permanently monitored. In a third stage, analytical redundancy is applied to monitor the sensors during the steady-state operation of the vehicle. Here, a vehicle model is used to check that the relationships between the sensor signals, as determined by vehicle motion, are plausible. These models are also frequently applied to calculate and compensate for sensor offsets as long as they stay within the sensor specifications.

In case of error, the system is switched off either partially or completely depending on the type of error concerned. The system's response to errors also depends on whether the control is activated or not.

Further names for driving-dynamics control

Aside from ESP® and the neutral designation ESC for Electronic Stability Control there are further names used by vehicle manufacturers for driving-dynamics control. Examples are Dynamic Stability Control (DSC), Vehicle Stability Assist (VSA), Vehicle Stability Control (VSC), Dynamic Stability and Traction Control (DSTC), Controllo Stabilità e Trazione (CST).

Special driving-dynamics control system for commercial vehicles

Function

Heavy commercial vehicles essentially differ from passenger cars in their much greater mass combined with higher centers of gravity and in additional degrees of freedom resulting from trailer operation [7]. They can thus assume unstable states that extend far beyond the skidding familiar to passenger cars. Such states include, as well as the jackknifing of multiple-stage vehicle combinations, caused for example by trailer sliding, overturning caused by high lateral acceleration. A vehicle-dynamics control system for commercial vehicles must therefore, in addition to providing the stabilization functions familiar to passenger cars, also prevent jackknifing and overturning.

Requirements

The following requirements, in addition to those for passenger cars, can be derived from the extended functions of driving-dynamics control for commercial vehicles:

- Improvement in directional stability and response of a vehicle combination (e.g. articulated road train or articulated-train combination) at physical driving limits in all operating and laden states. This includes preventing jackknifing on vehicle combinations.
- Reduced risk of overturning for a vehicle or vehicle combination in both quasi-stationary and dynamic vehicle maneuvers.

These requirements implemented in commercial-vehicle ESC lead, as is the case with passenger cars, to a significant improvement in driving safety. For this reason, European law will from 2011 require the gradual introduction of a driving-dynamics control system for heavy commercial vehicles (from 7.5 t) (integral part of ECE-R 13 [11]).

Application
In the meantime, commercial-vehicle ESC has become available for virtually all vehicle configurations (except all-wheel-drive vehicles):
– vehicles with the wheel formulas 4×2, 6×2, 6×4 and 8×4,
– combinations of tractor unit and semitrailer (articulated vehicle or simply semitrailer unit),
– combinations of rig and drawbar trailer (articulated road train),
– multiple trailer combinations (Eurokombi), e.g. combinations of rig, dolly and semitrailer or semitrailer unit with additional center-axle trailer or tractor unit with B-link and semitrailer.

Operating principle
Driving-dynamics control for commercial vehicles can be divided according to requirements into the two function groups described in the following.

Stabilizing the vehicle in the event of imminent skidding or jackknifing
Directional stabilization of a commercial vehicle is initially performed according to the same principles as for a passenger car. The controller compares current vehicle motion with vehicle motion desired by the driver, taking into account physical driving limits. The physical model of motion in the horizontal plane – for a single vehicle characterized by three degrees of freedom (longitudinal, lateral and yaw motion) – is however extended for an articulated vehicle to include the articulation angle between the rig and the trailer (an additional degree of freedom). There are further degrees of freedom involved for combinations with fifthwheel trailers.

To calculate the vehicle motion desired by the driver, the ECU uses simplified mathematical-physical models (single-track vehicle model, [8]) to determine the nominal yaw velocity of the rig. The parameters that are encountered in these models (characteristic vehicle speed v_{ch}, wheelbase l and steering ratio i_L) are either parameterized at the end of the vehicle assembly line or adapted to the vehicle's behavior during vehicle operation with the aid of special adaption algorithms (e.g. Kalman filters or recursive least-squares estimators, [9]). "Online" adaptation of parameters is particularly important in commercial vehicles, because the variety of variants and loads is much greater than for passenger cars.

Parallel to this, ESC calculates current vehicle motion from the measured variables available for yaw rate and lateral acceleration plus the wheel speeds. A significant deviation between the current vehicle motion and the motion expected by the driver leads to a control fault that is transformed by the actual controller into a corrective nominal yaw moment.

The level of the nominal yaw moment for a commercial vehicle depends on the control fault, the current vehicle configuration (wheelbase, number of axles, operation with or without trailer etc.) and the laden state (mass, center of gravity in the linear direction, etc.). As these parameters are variable, they are continuously determined by ESC. This is achieved, for example, in the laden state with the aid of an estimation algorithm that uses the signals from the engine management (engine speed and engine torque) and the vehicle linear motion (wheel speeds) to permanently identify the current vehicle mass.

On the basis of the current driving situation, the nominal yaw moment by braking individual or several wheels and of the trailer is transformed in a suitable manner. This is depicted by way of example in Figure 11a and in Figure 11b for a clearly defined oversteer and understeer situations respectively.

In addition to these clearly defined situations, there are other critical dynamic situations in which other wheels or wheel combinations are braked depending on the desired stabilization effect. Thus, for example in the case of sharper understeer, the entire vehicle is braked along similar lines to Enhanced Understeering Control (EUC) in passenger cars.

Because of the high center of gravity of commercial vehicles, skidding and jackknifing by such vehicles primarily occur at low-to-medium coefficients of friction at which the tire static-friction limit is already exceeded at an early stage. At high coefficients of friction, laden commercial vehicles, on account of their high center

of gravity, normally start to overturn before the static-friction limit of the tires is reached.

Reducing the risk of overturning
The overturning limit (lateral-acceleration limit) of a vehicle depends not only on the height of the center of gravity but also on the chassis systems (axle suspension, stabilizers, springs, etc.) and the type of payload (fixed or moving) [10].

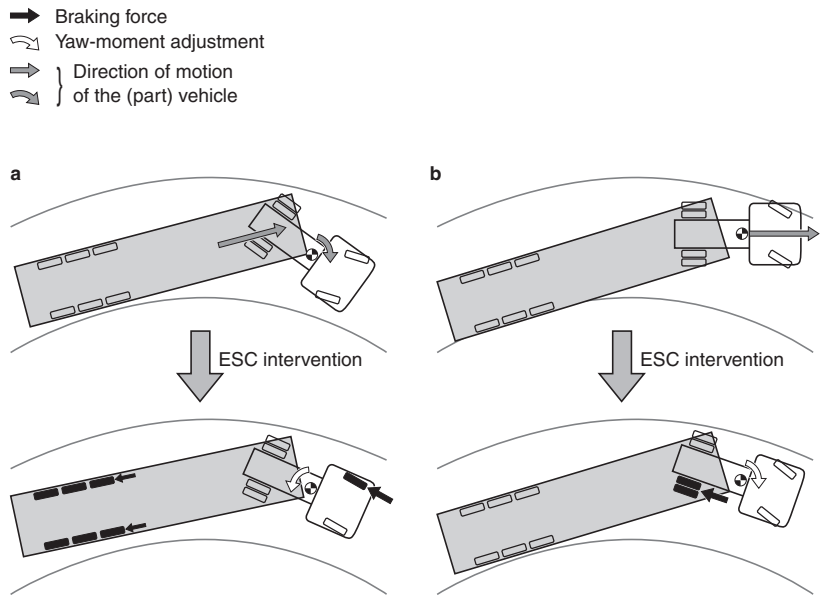
The situation which causes a commercial vehicle to overturn is, aside from a relatively low overturning limit, an excessive cornering speed. ESC makes use of this scenario to reduce the probability of the vehicle overturning. As soon as the vehicle approaches the overturning limit, it is slowed down by reducing engine torque and, if necessary, also applying the brakes. The overturning limit is determined here depending on the load of the vehicle and the load distribution, whereby the laden state of the vehicle is estimated "online".

Depending on each driving situation, the overturning limit that is determined is modified. Thus, the overturning limit in high-speed dynamic situations (e.g. obstacle-avoidance maneuvers) is reduced in order to permit early intervention. In very slow maneuvers, on the other hand (e.g. negotiating tight hairpin bends on uphill stretches), it is increased in order to prevent unnecessary and disruptive ESC intervention.

Determining the overturning limit is based on various assumptions regarding the height of the center of gravity and the dynamic response of the vehicle combination with a known axle-load distribution. This covers the largest portion of the usual vehicle combinations.

In order to ensure stabilization even in the case of strong deviations from these assumptions (e.g. extremely high centers of gravity), ESC additionally detects the lift of the wheels on the inside of a bend. This is achieved by monitoring the wheels

Figure 11: Principle of ESC braking intervention on semitrailers
a) Vehicle oversteering,
b) Vehicle understeering.



for implausible rotation speed. If necessary, the entire vehicle combination is then heavily decelerated by brake intervention.

A trailer wheel lifting on the inside of a bend is indicated by the trailer's electronically controlled braking system (ELB) via the CAN communication line (ISO 11992 [13]) by activating the ABS controller. For combinations with trailers equipped with ABS only, wheel-lift detection on the inside of a bend is limited to the tractor unit (rig).

System design

On the European market, the electronically controlled braking system ELB has come to the fore as the standard for brake control in heavy commercial vehicles. ESC is based on this system, extending it to include regulation of the driving dynamics. To do so, ESC uses the ELB capability of generating varying braking forces for each individual wheel independently of driver action.

The very different general conditions for commercial vehicle brake systems in North America mean that purely ABS or ABS/TCS systems are used as standard. An ESC based on ABS/TCS is therefore used for these and similar markets. Here, ESC uses the method already applied with TCS on the drive axle to generate braking force individually for each wheel independently of the driver by means of a TCS valve and the downstream ABS valves. In addition, for ABS-based ESC, the driver brake command must be measured by means of pressure sensors, which would otherwise not be possible during an ESC intervention on account of the function of the TCS valve.

Sensor systems

Like passengers cars, commercial vehicles use a combined yaw-rate and lateral-acceleration sensor and a steering-wheel-angle sensor as driving-dynamics sensors for ESC. Each of these sensors contains a microcontroller with a CAN interface for analyzing and safely transmitting the measured data.

The steering-wheel-angle sensor is usually mounted immediately below the steering wheel and it measures the angle of rotation of the steering wheel. This is then converted in the ECU into a wheel steering angle.

In order to pick up the lateral acceleration as close as possible to the center of gravity of the rig, the combined yaw-rate and lateral-acceleration sensor is usually mounted in the vicinity of the center of gravity.

Even though commercial vehicles essentially use the same sensors as passenger cars, the yaw-rate and lateral-acceleration sensor must have a much more robust design to cope with the rougher ambient conditions, particularly on the commercial-vehicle frame.

Electronic control unit (ECU)

The ESC algorithms are run together with the other algorithms for brake control (e. g. ABS and TCS) in the brake control unit. This control unit is constructed using conventional circuit-board technology with correspondingly powerful microcontrollers.

A CAN bus connects the ESC sensors with the control unit. The nominal brake pressures and wheel slip values of the ESC are then implemented by the relevant braking system for each wheel and for the trailer. In addition, the braking system transmits the requested engine torque via the vehicle CAN bus (usually standardized as per SAE J 1939 [14]) to the engine ECU for implementation.

Moreover, relevant information is also transferred from the engine and retarder to the braking system via the vehicle CAN bus. Essentially, this involves current and requested engine torque and speed, retarder torque, vehicle speed, and information from various control switches and any trailer that may be coupled.

Safety and monitoring functions

The extensive possibilities for ESC intervention in the handling characteristics of the vehicle and vehicle combination require a comprehensive safety system to ensure proper system functioning. This extends not only to the basic ELB or ABS/TCS system respectively, but also to the additional ESC components, including all sensors, ECUs and interfaces.

The monitoring functions used for ESC are essentially based on the functions used in passenger cars and are adapted to the characteristics of commercial vehicles.

There is also mutual monitoring of the microcontrollers distributed in the overall system. This means that the brake control unit contains a main computer and a monitoring computer that mainly performs plausibility checks alongside minor functional tasks. Furthermore, the corresponding algorithms permanently check memory and other internal computer hardware components to detect any defects that occur at an early stage.

The occurrence of faults results, depending on the nature and significance of the fault, in the shutdown of individual function groups through to a complete switching to "backup mode", in which the brakes are controlled by purely pneumatic means (fail-silent response). This ensures that incorrect sensor signals cannot cause implausible and possibly dangerous operating states in braking operations.

The occurrence of a fault is indicated to the driver by suitable means (e.g. warning lamp or display) so that suitable action can be taken.

Furthermore, any faults that occur are assigned a time stamp in the control unit and stored in the fault memory. The workshop can evaluate these with the help of a suitable diagnosis system.

References

[1] E.K. Liebemann, K. Meder, J. Schuh, G. Nenninger: Safety and Performance Enhancement: The Bosch Electronic Stability Control. SAE Paper Number 2004-21-0060.

[2] National Highway Traffic Safety Administration (NHTSA) FMVSS 126: Federal Motor Vehicle Safety Standards; Electronic Stability Control Systems; Controls and Display. Vol. 72, No. 66, April 6, 2007.

[3] M. Mitschke, H. Wallentowitz: Dynamik der Kraftfahrzeuge. 4th Edition, Springer-Verlag, 2004.

[4] A. van Zanten, R. Erhardt, G. Pfaff: FDR – Die Fahrdynamikregelung von Bosch. ATZ Automobiltechnische Zeitschrift 96 (1994), Volume 11.

[5] A. Trächtler: Integrierte Fahrdynamikregelung mit ESP®, aktiver Lenkung und aktivem Fahrwerk. at – Automatisierungstechnik 53 (1/2005).

[6] K. Reif: Automobilelektronik, 3rd Edition, Vieweg+Teubner, 2009.

[7] E. Hoepke, S. Breuer (Editors): Nutzfahrzeugtechnik – Grundlagen, Systeme, Komponenten. 4th Edition; Vieweg Verlag, 2006.

[8] C.B. Winkler: Simplified Analysis of the Steady State Turning of Complex Vehicles. International Journal of Vehicle Mechanics and Mobility, 1996.

[9] Ali H. Sayed: Adaptive Filters. John Wiley & Sons, 2008.

[10] D. Odenthal: Ein robustes Fahrdynamik-Regelungskonzept für die Kippvermeidung von Kraftfahrzeugen. Dissertation TU München, 2002.

[11] ECE-R 13: Uniform provisions concerning the approval of category M, N and O vehicles with regard to brakes.

[12] ECE-R 13-H: Standard conditions for approval of passenger cars with regard to brakes.

[13] ISO 11992: Road vehicles – Interchange of digital information on electrical connections between towing and towed vehicles.

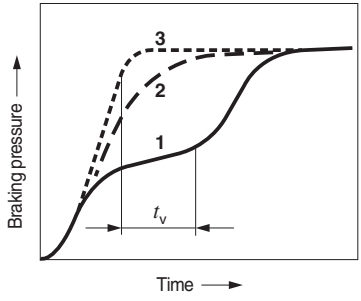
[14] SAE J 1939: Serial Control and Communications Heavy Duty Vehicle Network Top Level Document.

Automatic brake-system operations

The supplementary functions described here utilize except for the pneumatic-mechanical brake assistant the driving-dynamics control infrastructure. This comprises the sensors, the actuators and the ECU.

Brake-assistant operations
Investigations into driver braking behavior in the 1990s showed that car drivers differ in the way that they respond to braking situations. The majority – the "average drivers" – do not brake hard enough when faced with an emergency situation, in other words, they require an unnecessarily long braking distance (Figure 1). This can be remedied by a system first launched in 1995, the brake assistant. Its main purposes are as follows:
– It interprets a certain rate of pedal movement (rapid application of the brakes) that fails to apply maximum braking force as an intention by the driver to carry out full braking. In this case, it generates the braking pressure required to achieve full braking effect.
– It allows the driver to "cancel" full braking operation at any time.

Figure 1: Comparison between braking with and without brake-assistant operations
A longer braking distance is required without a brake assistant than with a brake assistant.
1 "Average driver",
2 Experienced driver,
3 "Average driver" with brake assistant.
t_v Delay time for braking results in longer braking distance.



– The behavior of the brake booster and, therefore, the pedal feedback, is not altered under normal braking conditions.
– The basic braking-system function is not diminished if the brake assistant fails.
– The system is designed to prevent accidental activation.

Pneumatic brake assistant
This system requires a modified brake booster which increases the level of amplification according to the rate of pedal movement and pedal force. This results in faster and greater buildup of pressure in the wheel brakes.

In an alternative version, the brake booster is extended to include an electronically activated valve. This enables an ECU to influence the pressure difference between the brake-booster chambers and thus the boosting of the braking force. This provides better opportunities for optimizing the trigger threshold and response characteristics.

Hydraulic brake assistant
The hydraulic brake-assistant operation uses the driving-dynamics control hardware. A pressure sensor detects the driver's braking intention; the ECU analyzes the signal on the basis of the defined triggering criteria, and initiates an appropriate brake-pressure build-up in the hydraulic system. The upstream brake booster is a standard unit and does not require any modification.

As a general observation, it should be stated that it is an absolute requirement that all brake-assistant system variants referred to are used in conjunction with an antilock braking system (ABS) or a driving-dynamics control system due to the actively generated rapid brake-pressure rise beyond the wheel-lock limit.

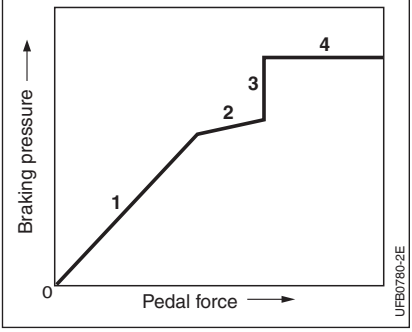
Automatic brake-pressure increase at the rear wheels

This is a function which provides the driver with additional brake servo assistance for the rear wheels if the front wheels are controlled by the ABS system. This function was introduced because many drivers do not increase the pedal force at the start of ABS control even though the situation would require this. Once ABS control has been initiated on the front wheels, the wheel pressures at the wheels on the rear axle are increased by way of the return pump of the hydraulic modulator until these too have reached the lockup-pressure level and ABS control is initiated (Figure 2). The brake application is therefore at the physical optimum. The pressure in the rear axle wheel-brake cylinders can then exceed the pressure in the brake master cylinder, also during ABS control.

The cutoff condition is fulfilled when the wheels on the front axle are no longer under ABS control or when the pressure in the brake master cylinder falls below the cutoff threshold.

Figure 2: Automatic brake-pressure increase

1 Brake-pressure increase by brake booster,
2 Further brake-force increase by pedal force,
3 Brake-pressure increase by driving-dynamics control hydraulic modulator,
4 ABS control range.



Automatic brake-pressure increase with forceful pressing of brake pedal

This function provides the driver with additional brake servo assistance. It is activated if the maximum possible vehicle deceleration is not achieved even if the driver forcefully presses the brake pedal to the point that would normally cause the lockup pressure to be reached (primary pressure over approx. 80 bar). This is the case, for example, at high brake-disk temperatures or if the brake pads have a considerably reduced coefficient of friction.

When this function is activated, the wheel pressures are increased until all wheels have reached the lockup pressure level and ABS control is initiated (Figure 2). The brake application is therefore at the physical optimum. The pressure in the wheel-brake cylinders can then exceed the pressure in the brake master cylinder, also during ABS control.

If the driver reduces the desired level of braking to a value below a particular threshold value, the vehicle deceleration is reduced in accordance with the force applied to the brake pedal. The driver can therefore precisely modulate the vehicle deceleration when the braking situation has passed. The cutoff condition is fulfilled if the primary pressure or vehicle speed falls below the respective cutoff threshold.

Brake-disk wiper

This function ensures that, in the event of rain or a wet road, the splash water is cyclically removed from the brake disks. This is achieved by automatically setting a low brake pressure at the wheel brakes. In this way, the function helps to ensure minimum brake-response times when driving in wet conditions. Evaluated signals from the windshield wipers or rain sensor are used to detect wet conditions.

The pressure level is adjusted so that the vehicle deceleration cannot be perceived by the driver. Actuation is repeated at a defined interval for as long as the system detects rain or a wet road. If required, just the disks at the front axle can be wiped. The wiping procedure is terminated as soon as the driver applies the brakes.

Automatic prefill
This function reduces the total braking distance in emergency situations whereby the driver operates the brake pedal immediately after the accelerator pedal is released. This is achieved by prefilling and thereby preloading the brake system after release of the accelerator pedal, which means that in the subsequent brake operation the pressure build-up is considerably more dynamic. Accordingly, high vehicle deceleration sets in earlier.

Prefilling of the brake system is adjusted by the return pump of the hydraulic modulator. The brake shoes are then firmly applied to the brake disks. If there is no operation of the brake pedal directly after a rapid release of the accelerator, the pressure in the brake system is reduced again. This does not impair driveability.

Electromechanical parking brake
The electromechanical parking brake (EMP) generates the force for applying the parking brake by electromechanical means. The function of the hand- or foot-operated parking-brake lever is performed by a control knob with an electric-motor-and-transmission combination. When the driver operates the control knob, the electric motor (actuator) is activated when the system detects that the vehicle is at a standstill. When the vehicle is parked on a level surface, the holding forces are set lower than when the vehicle is fully laden and parked on a gradient. Active wheel sensors are used to detect that the vehicle is parked (i.e. stationary). As an option, it is also possible for the road gradient to be detected by a tilt sensor.

The parking brake is released by means of the same control knob. However, various safety regulations and requirements have to be met, e.g. to prevent inappropriate or inadvertent release of the parking brake by children or animals (see also Parking brake, Passenger-car brake system).

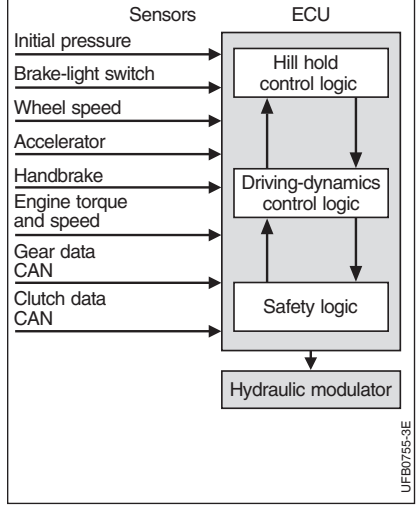
Controlled braking with hydraulic modulator
If the electromechanical parking brake is activated while driving, the vehicle must be safely decelerated to a stop. The brake pressure required for this is built up by the return pump of the hydraulic modulator. The driving-dynamics control system ensures safe braking even on slippery or wet road surfaces. When the vehicle has come to a stop, the electromechanical parking brake assumes the function of keeping the vehicle stationary.

The driver must press the activation knob of the automatic parking brake continuously during the deceleration phase.

Hill hold control
This system simplifies hill starts. It prevents the vehicle from rolling back after the driver has released the brake pedal. This is particularly helpful on heavily laden vehicles with manual transmission and vehicles that are towing trailers. There is no need to operate the parking brake. The function also works when performing a hill start in the reverse direction.

The system detects the driver's intention to pull away (Figure 3). After the brake pedal is released, there is approximately two seconds to start pulling away. The

Figure 3: Hill hold control



brake is released automatically here if the drive torque is greater than the torque resulting from downgrade force.

Low-speed traction control is based on the driving-dynamics control hardware and additional sensors: a tilt sensor detects the road gradient, a gear switch detects whether the driver has shifted to reverse gear, and a clutch switch recognizes whether the driver has depressed the clutch pedal.

Automatic braking on hill descent – hill descent control

This system is a convenience function which assists the driver on offroad descents with gradients of approximately 8 to 50 % by automatically operating the brakes. The driver can then concentrate fully on steering the vehicle and is not distracted by the need to operate the brakes at the same time. The brake pedal does not have to be operated.

When this system is activated, e.g, by pressing a button or switch, a preset speed is maintained over the extent of the specified brake pressure. If required, the driver can vary the predetermined speed by pressing the brake and accelerator pedal or using the control buttons of a speed control system.

The system remains active until it is switched off by pressing the button or switch again, i.e. it is not automatically deactivated.

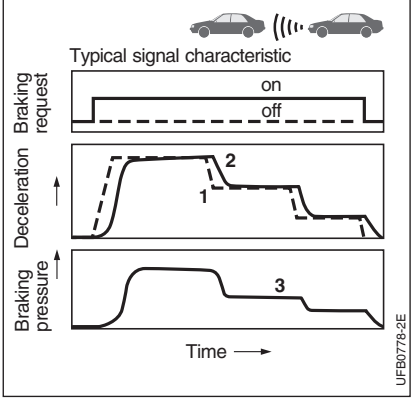
Automatic brake application for driverassistance systems

This function is an additional function for active brake application with adaptive cruise control (ACC), i.e. for automatic vehicle-to-vehicle ranging. The brakes are applied automatically without the driver pressing the brake pedal as soon as the distance to the vehicle in front falls below a predetermined distance (Figure 4). This is based on a hydraulic brake system and driving-dynamics control.

The function receives a request to decelerate the vehicle by a desired amount (input). This is calculated by the upstream ACC system. Automatic brake intervention maintains vehicle deceleration by means of appropriate brake pressures which are adjusted with the aid of the hydraulic modulator.

Figure 4: Function description of automatic brake application for ACC

1 Deceleration specified by ACC (Adaptive Cruise Control),
2 Current deceleration,
3 Pressure in wheel-brake cylinders.



Motorcycle Stability Control

Application

Motorcycle Stability Control (MSC) improves two-wheeler safety in all critical riding situations – particularly in turns/curves. Here MSC coordinates all the functions relevant to riding dynamics such as, for example, the Antilock Braking System (ABS, see Figure 1), the electronically Combined Brake System (eCBS), anti-stoppie control, Motorcycle Traction Control (MTC), and wheelie control taking account of the current riding situation and the current road friction-coefficient conditions.

The ABS riding-dynamics control of a motorcycle is similar in some points to that of a passenger car (see Wheel-slip control systems). Therefore the riding-dynamic particulars of two-wheelers will be discussed especially in the following.

Riding physics of a two-wheeler

Maintaining equilibrium

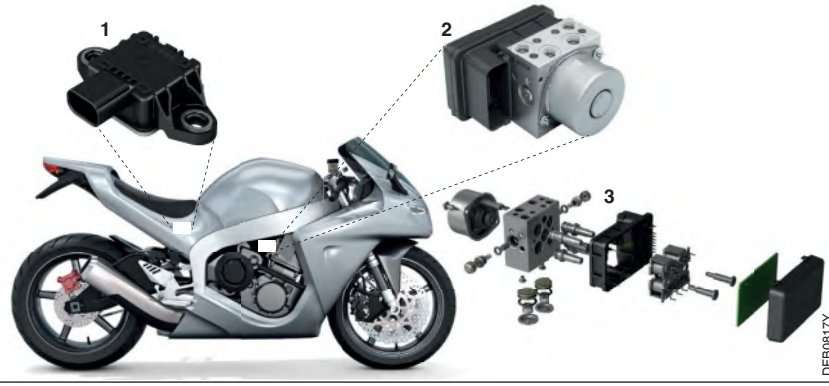
A two-wheel vehicle is in a state of unstable equilibrium and is prone to falling over. When moving it is 100% dynamically stabilized by the gyroscopic forces of the turning wheels [1] and via the castor offset. A moving two-wheeler thus becomes stable from a certain speed.

Because a two-wheeler can tilt about the longitudinal axis, it must maintain equilibrium. Figure 2 shows the equilibrium of forces between gravity and centrifugal force when cornering. In the case of a real tire the wheel contact point is laterally offset. As a result, more inclination through the angle λ is required than for a very thin tire. The wider the tire, the greater the necessary inclination at the same speed and for the same curve radius.

Self-stabilization

If the motorcycle tilts towards the inside of the curve, gyroscopic effects on the turning front wheel induce a steering moment in the direction of the inclination [1]. As a result the motorcycle rides a tighter radius, whereupon the rising centrifugal

Figure 1: MSC vehicle components
1 Inertial sensor unit,
2 ABS hydraulics with ECU (assembly)
3 ABS hydraulics (exploded view).



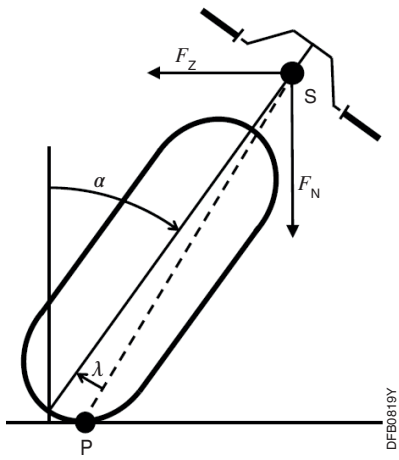
force rights the motorcycle again. The righting motion turns the front wheel back to the initial position. The castor offset on the front wheel brings about a similar effect. The distance between the point of intersection of the steering axis with the road and the wheel contact point is called the castor offset. When the lateral component of the wheel downforce increases, this results in a corresponding torque towards the inside of the curve on the steering (see Figure 3). The stability of a two-wheeler increases with the speed and the resulting increasing gyroscopic effects. Likewise, a greater castor offset provides for better straight-ahead stability.

To ride a stationary curve, the rider must apply a steering moment to the outside of the curve to compensate the steering moment caused by gyroscopic effects and castor offset. If the rider lets go of the handlebars, the two-wheeler rights itself and rides straight ahead again.

This produces the decisive difference as the rider specifies the desired curve radius. In a passenger car the steering wheel is turned without greater consideration of the steering torque to the corresponding steering angle. In contrast, the rider on a two-wheeler specifies the curve radius via the steering moment and the two-wheeler adjusts the associated steering angle itself accordingly by means of self-stabilization. Here the two-wheeler and the steering angle alway oscillate about the position of equilibrium somewhat. In addition, as determined by the geometry, the self-adjusting steering angle decreases as the inclination increases.

Figure 2: Equilibrium of forces between gravity and centrifugal force when cornering

S Center of gravity,
P Wheel contact point,
F_Z Centrifugal force,
F_N Gravity (normal force),
α Inclination angle,
λ Angle to tire center plane.



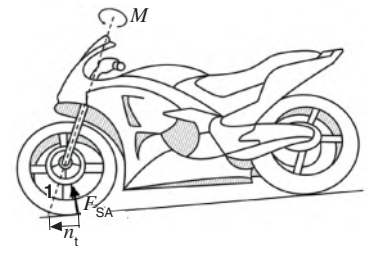
Inclination estimation

Important riding-dynamic variables of a two-wheeler are the inclination (roll angle) and the pitch angle because a motorcycle leans far into the curve and tends to roll over when subjected to heavy braking and acceleration. To determine these two angles an inertial sensor unit with three acceleration sensors and three yaw sensors was developed (Figure 1); this sensor unit enables changes in angle to be sensed quickly.

The inclination and the pitch angle are calculated in accordance with ISO 8855 [2] from the direction in which Earth's gravity points when viewed from the vehicle. The direction components of the gravity vector $g = (g_x, g_y, g_z)$ are calculated by means of the following two methods:

Figure 3: Self-stabilization via the castor offset of a motorcycle leaning to the right

1 Steering axis.
M Steering moment,
F_{SA} Lateral component of wheel downforce,
n_t Castor offset.



1. Determination of the gravity vector from accelerations
When the vehicle is stationary the three acceleration sensors directly measure the components of the acceleration due to gravity. When the vehicle is moving, further components brought on by the change of speed and the centrifugal forces are added. The measured accelerations must be corrected by these components in order to obtain the gravity vector. Because the acceleration signals are for the most part disrupted more strongly by engine vibrations or road irregularities, the vector obtained in this way $g = (g_x, g_y, g_z)$ must be filtered with a low-pass filter to determine the angle precisely. This worsens the signal dynamics, but the value is long-term stable for this.

2. Rotation of the gravity vector by integration of the measured rotational movement
The rotational movement of the motorcycle about each axis is measured by three yaw sensors in the form of angular velocities. From this the change of direction of $g = (g_x, g_y, g_z)$ is calculated in each calculation cycle. The direction vector obtained in this way has good dynamics because it follows the yaw rates without further filtering. However, measuring errors or disturbances are also added up. As a result, when considered over a longer period of time, the direction of the vector would deviate ever further from the real Earth's gravity.

Correcting the g vector from method 2 with that from method 1 combines the advantages of both methods. The resulting angles dynamically follow each change and are at the same time also long-term stable.

Motorcycle ABS

Many accidents resulting in serious injury could be avoided if the rider braked fully without decelerating prior to the collision. MSC was therefore developed with the objective of enabling the rider to perform drastic braking in every situation without resulting in an accident.

Function of ABS
Just like in a passenger car, the antilock braking system (ABS) is intended to stop the wheels from locking. On a motorcycle particular attention is paid to the front wheel because here the greatest braking force can be applied and at the same time the front wheel is crucial to riding stability.

Usually the wheel slip is the most important variable for the ABS controller. However, the vehicle speed must be determined for this purpose. This poses a particular challenge when it comes to two-wheelers. Thus, for example, the rear wheel is lifted up in the event of a “stoppie” in response to heavy braking with the result that the rear-wheel speed no longer matches the driving speed. From the remaining front-wheel speed the vehicle speed must simultaneously be estimated and the wheel slip calculated. Because both these variables can no longer be reliably determined in such a situation, motorcycle ABS uses as well as the slip additional indicators, such as for example wheel deceleration, to detect wheel instability.

Rear-wheel lift-up control
A rider who brakes heavily on two wheels decelerates primarily with the front wheel. Due to the dynamic wheel-load distribution in combination with a high center of gravity, the wheel load shifts sharply towards the front wheel – often even resulting in the rear end lifting up. Rear-wheel lift-up control was developed to make stable braking possible. This function evaluates the slip and deceleration values of both wheels. If the rear wheel threatens to lift up while braking, the function adapts the pressure in the front-wheel brake circuit to keep the rear wheel on the ground and avoid rollovers.

Using the progression of the rear-wheel speed, it is in fact possible to detect when the rear wheel has lifted up, however this process does not distinguish whether the wheel hovers just a few millimeters above the ground or whether a rollover is already imminent. Here the pitch-angle and pitch-rate information from the inertial sensor helps to detect the situation and intervene with optimal control.

Braking in a turn/curve

Normal motorcycle ABS is optimized to as short a braking distance as possible during straight-ahead riding. The high brake-slip values necessary for this purpose could cause a side-slip in a curve because the tire is no longer able to apply the side control force. If the tire slips, gravity from Figure 2 is predominant and the front wheel swings in due to the castor offset towards the inside of the curve, drastically increasing the angle between direction of travel and wheel. The resulting Friction between tire and road brakes the wheel to such an extent that it locks, even if the brake pressure were reduced in this situation to 0 bar. A fall is virtually inevitable. It is therefore essential when cornering not to exceed the critical slip value.

Steering moment under braking conditions

A second critical point in any braking at an inclination on the front wheel is the steering moment under braking conditions. Because the tire when inclined in no longer centrally touching the road, a longitudinally acting braking force brings about a steering moment which turns the wheel towards the inside of the curve (Figure 4). The tighter curve radius allows the centrifugal force to increase and the motorcycle rights itself and ultimately rides a wider radius than at the start. If the rider fails to correct this already at the initial-braking stage, the motorcycle may unintentionally stray onto the oncoming carriage way or leave the road.

Gradient control

Both unintentionally high brake slip in a curve and a sudden jump in the steering moment under braking conditions can be effectively brought under control by an inclination-dependent gradient control of the brake pressure. The slowly rising brake pressure gives the rider time to counteract the building steering moment under braking conditions. In addition, the slowing vehicle requires a certain righting moment because the motorcycle must be upright again at the latest when stationary.

With more slowly increasing brake pressure the brake slip as well builds up under controllable conditions and does not overshoot the limit range.

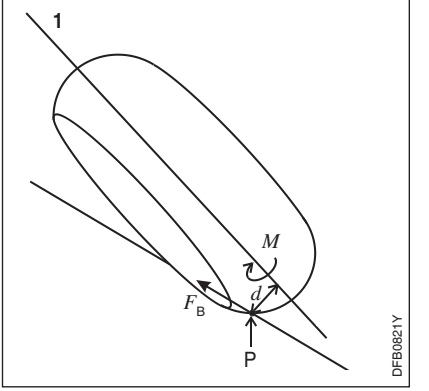
A proficient rider starts with a low brake pressure when braking in a curve and then increases this with decreasing speed and inclination. Gradient control is optimally set in such a way that it is not perceived even by professional racing riders. Only in a drastic-braking situation does it show its potential.

Further adaptations in a curve

Similarly to passenger cars, the ABS control thresholds and control parameters are adapted on a curve-dependent basis in

Figure 4: Steering moment under braking conditions

1 Steering axis.
P Wheel contact point,
M Steering moment under braking conditions,
F_B Braking force,
d Distance to center.



order to achieve an optimal combination of stability and braking distance. Furthermore, high ABS brake-pressure modulations must be avoided at an inclination on the front wheel because these could lead to dangerous handlebar vibrations due to the steering moment under braking conditions.

Electronically controlled braking-force distribution

To improve riding safety particularly for heavy motorcycles, Bosch has developed "ABS enhanced". This facilitates a combined braking system in the form of electronically controlled braking-force distribution (eCBS, electronically Combined Brake System). Weighing just 1.5 kg and with a volume of 1.2 liters, it is the world's smallest and lightest ABS with eCBS.

Concept of the combined braking system

With a combined braking system both wheels can be braked, even though the rider only applies the front or the rear brake (Figure 5). Using the measurement of the rider brake pressure and further parameters such as for example the speed, the system calculates the optimal braking-force distribution in each braking situation. In phase 1 in Figure 5 the rider applies the brake, in phase 2 the nominal pressure on the other wheel is calculated according to the brake pressure specified by the rider, and in phase 3 brake pressure is finally built up on the other wheel by activating the valves.

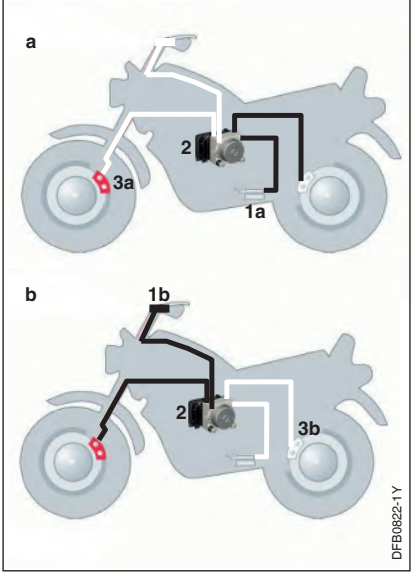
This enables very short braking distances to be achieved without a loss of stability, even if the rider, for example when panic braking, is only braking with the rear wheel. Optimization of braking-force distribution likewise reduces pitching and smooths abrupt vehicle movements, allowing the physical limits to be exploited to better effect. This makes "ABS enhanced" a safe and comfortable motorcycle-braking system.

Operating principle

The rider brake pressure is measured at the master cylinder for the eCBS function. According to the vehicle-specific braking-force distribution characteristic the systems calculates the target pressure at the other, unbraked wheel and actively builds up the brake pressure there by means of a pump. Because the target pressure is built up electronically, the brake pressure is not as in conventional, mechanical systems limited by the rider pressure. This electro-hydraulic approach offers a higher degree of freedom in the optimization of braking-force distribution to the specific riding situation. Thus, for example depending on the rear-wheel pressure, more brake pressure can be built up at the front at high speed than at low speed. Furthermore, if a speed threshold is undershot, the eCBS function can be deactivated completely so as to

Figure 5: Combined braking system – basic concept and purpose

a) Electronic pressure build-up at the front brake caliper,
b) Electronic pressure build-up at the rear brake caliper.
1a Rider applies rear brake,
1b Rider applies front brake,
2 the nominal pressure to the other wheel is calculated according to the rider brake pressure,
3a Brake pressure is built up at the front wheel by activating the valves and the pump,
3a Brake pressure is built up at the rear wheel by activating the valves and the pump.



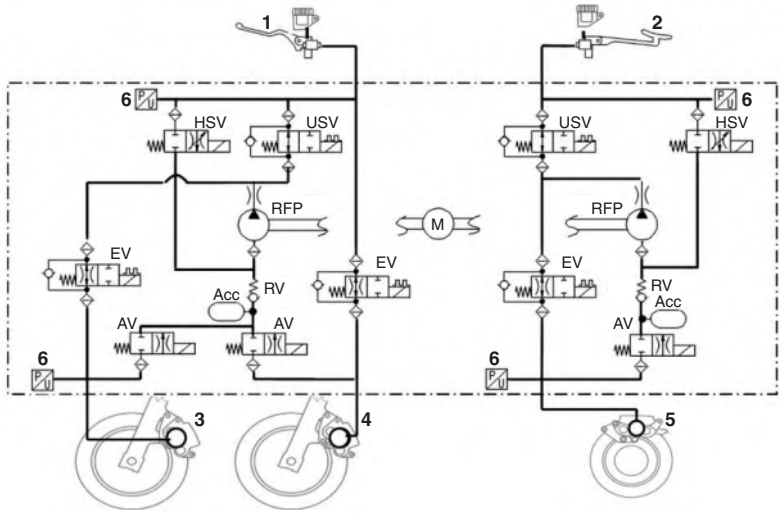
avoid the comfort problems familiar from conventional systems when braking in a curve at low speed. Not least, in contrast to conventional systems, no additional hardware components are required – the complete eCBS function is already featured in “ABS enhanced”.

System design of the combined braking system
The combined braking system is available as a 2- and a 3-channel version. In the 2-channel version there is a separate brake circuit for the front wheel and one for the rear wheel. Both are independent of one another. Thus, for example for the active build-up of pressure at the front no brake fluid from the rear brake circuit is used and vice versa. In addition to the inlet (IV) and outlet valves (OV), which use a basic ABS, two additional valves per brake circuit are required – a switchover valve (SV) and a high-pressure switching valve (HSV).

The technical concept for the 2-channel system becomes apparent from Figure 6 (the figure shows a 3-channel hydraulic system, the third channel is explained later): When the rider brakes at the rear, the target pressure at the front is calculated for optimal braking-force distribution. For the active pressure build-up the high-pressure switching valve is opened and the suction pump delivers brake fluid. At the same time, energizing the changeover valve regulates its throughflow, which adjusts the desired pressure difference between master cylinder and brake caliper.

“ABS enhanced” as a 3-channel system
When the rider simultaneously brakes at front and rear, the 2-channel system has certain disadvantages with regard to the handle feel:
– If only the rear brake is applied, the system also brakes at the front wheel. If now the rider also brakes at the front,

Figure 6: Combined braking system – 3-channel hydraulic schematic
1 Brake handle, 2 Brake pedal,
3 Brake caliper 1 On front wheel, 4 Brake caliper 2 On front wheel,
5 Brake caliper on rear wheel, 6 Pressure sensors.
IV Inlet valve, OV Outlet valve,
HSV High-pressure switching valve, SV Switchover valve,
RP Return pump (suction pump), M Pump motor,
NRV Non-return valve, Acc Accumulator volume.



the brake handle feels hard. This could irritate the rider.
– If the rider is already braking lightly at the front and then applies the rear brake, the combined braking system builds up additional brake pressure at the front. In this case the connection between the brake-fluid reservoir and the brake line is already closed. The volume displacement required for the pressure build-up leads to the brake handle being drawn in.

Although both effects also occur at the rear brake circuit, the influence there on braking comfort is less pronounced because the sensitivity at the pedal is lower.

One way of solving the comfort problems of the 2-channel system is to equip the front brake with two calipers, with each being supplied by its own channel. The capacity for active pressure build-up is only implemented for one of the two calipers. When the combined braking system builds up pressure on the front wheel when the brake pedal is pressed, one caliper remains depressurized. Figure 6 shows the associated hydraulic schematic of the 3-channel system.
– If the rider now additionally applies the brake, the previously unactuated caliper through its elasticity provides for a natural braking feel.
– Due to the fact that, as shown in Figure 6, pressure is only actively built up at the left caliper of the front wheel, less brake-fluid volume is needed to achieve a particular deceleration (because for example there is only clearance at one caliper). As a result the handle reaction is reduced to an acceptable level.

Motorcycle Traction Control

Traction control regulates traction slip at the driven wheel by reducing the drive torque. This can prevent the rear wheel from breaking away sideways in a curve and optimizes the transmission of traction force to the road surface. Because the radii, the contours, and the widths of motorcycle tires are different on the front and rear wheels, the effective tire rolling circumferences depend to varying degrees on the inclination. Consequently, knowledge of the inclination angle brings a decisive advantage when calculating the wheel-slip values and in the performance capability of traction control.

Above all for Enduro motorcycles there is from slick through road tires right down to massive-bar offroad tires a wide range of possible tire combinations which can be used on all manner of surfaces such as dry asphalt, cobblestone pavement, and sand. Therefore Motorcycle Traction Control permits for the most part a wide adjustment range for adaptation to meet demand.

Anti-wheelie control

Anti-wheelie control controls the engine torque in such a way as to prevent the front wheel from lifting up in an uncontrolled manner and at the same time to guaranteed the maximum possible acceleration.

Using the progression of the front-wheel speed, it is possible to detect when the front wheel has lifted up. However this process does not distinguish whether the wheel hovers just a few millimeters above the ground or whether a rollover is already imminent. Here the pitch-angle and pitch-rate information from the inertial sensor helps to detect the situation and intervene with optimal control.

References
[1] J. Stoffregen: Motorradtechnik. 9th Ed., Verlag Springer Vieweg, 2018.
[2] ISO 8855: Road vehicles – Vehicle dynamics and road-holding ability – Vocabulary.

Integrated driving-dynamics control systems

Overview

In addition to the brakes and engine, as used by the driving-dynamics control system (Electronic Stability Control, ESC), there are other actuators in the chassis and in the drivetrain which are used to purposefully influence driving dynamics. The functional combination of actuators relevant to driving dynamics is referred to by different names by the various automobile manufacturers and suppliers: Vehicle Dynamics Management (VDM), Integrated Chassis Control (ICC), Integrated Chassis Management (ICM), and Global Chassis Control (GCC).

Primary uses
The functions are used primarily in the fields of directional stability, agility and reducing "driver workload", i.e. the efforts exerted by the driver to operate and control the vehicle.

Improving directional stability
In critical driving situations, these functions take on stabilization tasks which are normally performed by the driver himself. Typical examples are the oversteer situation or braking on roads with different grip factors on the left and right sides of the vehicle (μ split).

Increasing agility
Some of these functions improve the acceleration response of the lateral-dynamic response of the vehicle to steering inputs by the driver. These make the vehicle more agile, i.e. it responds more spontaneously and dynamically to driver inputs.

Reducing the effort to control/operate the vehicle
An improved response by the vehicle to movements of the steering wheel and automatic stabilization interventions relieve the driver's workload, the steering effort in particular being reduced.

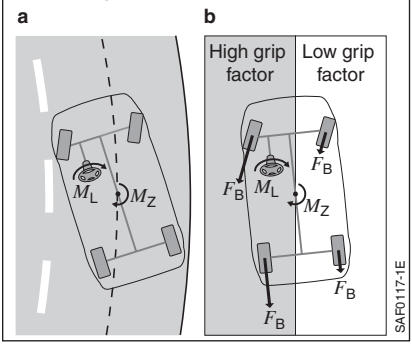
Functions

Driving-dynamic driver steering recommendation
Function
This function utilizes electrical power steering as the actuator. An additional, driving-dynamically motivated steering torque is superimposed on the steering-servo torque of the power steering in order to provide the drive with a steering recommendation.

Limitation of supplementary steering torque
The driving-dynamic supplementary steering torque is limited, depending on the vehicle, to values of approximately 3 Nm so that the driver can oversteer the steering recommendation at any time.

Driving-dynamics applications
The steering recommendation is activated in different driving situations. In a oversteer situation, a steering torque is introduced in the countersteer direction (Figure 1a). In the event of understeer, the function motivates the driver not to increase the steer

Figure 1: Driving-dynamic driver steering recommendation with supplementary steering torque
a) Oversteer,
b) Braking on road with different grip factors.
M_L Supplementary steering torque,
M_Z Yaw moment based on intended driver reaction,
F_B Braking force.



angle further, since the side-force potential on the front wheels is already exhausted and further steering may even reduce the side force.

When braking or accelerating on roads with different grip factors on the left and right sides of the vehicle (μ split), the driver is assisted in compensating the vehicle yaw motion by countersteering (Figure 1b).

Active steering stabilization on the front axle

Function

This function intervenes via the angle-superimposing steering actuator of the override steering directly in the vehicle movement. An additional steering angle adjusted by an electric motor is added to the driver's steering angle in an override gearbox.

Driving-dynamics applications

The function operates, for example, in an oversteering driving situation. Yaw-rate control with automatic steering-angle intervention returns the excessively high yaw-rate value to its setpoint value.

During braking on roads with different grip factors, the function provides for compensation of the yaw motion. Since yaw compensation is automatically triggered and happens much faster than the steering operation of a typical driver, the measures in driving-dynamics control for yaw-moment build-up delay can be reduced. The braking distance is also reduced.

A further benefit can be achieved in the event of understeer and starting on roads with different grip factors.

Demands on the sensor system

Because of the high positioning rate of an override steering system, the sensor signals must be monitored with low fault latencies, i.e. a fault must be signaled quickly. This means that the inertial sensor system must have a redundant design.

Active steering stabilization on the rear axle

Function

In vehicles with rear-axle steering, the primary function controls the rear-axle steering angle as a function of the steering-wheel angle and the driving speed. At low speeds, the rear wheels are steered in the opposite direction to the front wheels, which improves vehicle maneuverability.

At high speeds, the rear wheels are steered in the same direction as the front wheels. In this way, the yaw motion is less intensely excited during obstacle-avoidance maneuvers. This increases directional stability considerably.

Driving-dynamics applications

The associated functions act in situations similar to active front-axle steering interventions. The effectiveness of oversteer interventions is crucially dependent on the current utilization of the side-force potential on the rear wheels.

When braking on roads with different grip factors, it is necessary on account of the reduction of load on the rear axle to make interventions with larger angles in order to suppress the yaw motion. On the other hand, braking can take place without a considerable attitude angle.

Demands on the sensor system

The functions for rear-axle steering likewise require a redundant inertial sensor system, since the positioning rate of a steering actuator on the rear axle is comparable with the override steering.

Torque distribution in the longitudinal direction for four-wheel drives

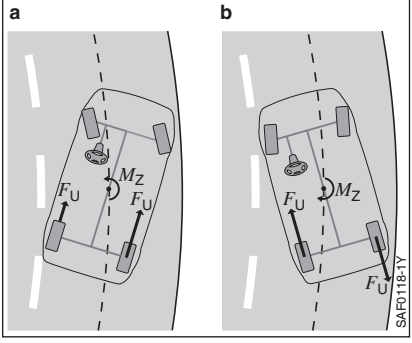
Function

In vehicles with four-wheel drive, interaxle differential locks or transfer cases within the framework of their basic function improve the traction and, with it, the accelerating performance of the vehicle, whereby the wheels of both axles transfer tractive forces.

Driving-dynamics applications
In situations where the grip factor is not fully utilized by traction, the self-steering properties can be altered by shifting the tractive forces between the front and rear axles. The more tractive force an axle transfer, the more the side-force potential on the wheels of this axle is weakened. Shifting the tractive force to the front axle increases the understeer tendency, while shifting it to the rear axle reduces the understeer tendency. Through dynamic control of drivetorque distribution, lateral-dynamic agility can be increased without the vehicle tending towards oversteer.

Wheel-torque distribution in the lateral direction
Function
Functions based on a "torque-vectoring actuator" make a significant contribution to increasing agility. The actuator permits an extensively free shift of the wheel torque between the left and right wheels on an axle. In this way, the total turning force excluding friction losses is not reduced in the actuator meaning that this intervention is virtually neutral with regard to the driving speed.

Figure 2: Wheel-torque distribution in the lateral direction
a) Understeer,
b) Oversteer.
F_U Wheel peripheral force (driving, braking),
M_Z Yaw moment based on intervention.



Driving-dynamics applications
In an understeering driving situation, the drive torque of the wheel on the outside of the curve is increased and the drive torque of the wheel on the inside of the curve is decreased. In this way, an additional yaw moment turning into the curve acts on the vehicle (Figure 2a). The understeer tendency decreases; the vehicle's lateral-dynamic agility increases.

When accelerating on roads with different grip factors, the drive torque is specifically directed to the wheel with a high grip factor. Brake interventions by the traction control system on the wheel with a low grip factor are dispensed with to a large extent. This increases the mean acceleration on roads with different grip factors.

Even in an oversteering situation, the brake interventions are partly replaced by a wheel-torque shift (Figure 2b). In this way, the speed loss can be reduced by brake interventions of the driving-dynamics control in the event of mild oversteer. In critical oversteering situations, the brakes still intervene, since in this case the speed loss is desired to alleviate the driving situation.

Replacement of wheel-torque distribution by brake and engine-control interventions
The understeer interventions by means of a torquevectoring actuator can be emulated by brake interventions on the wheels on the inside of the curve. The engine torque is increased to compensate the deceleration by the one-sided brake intervention. This produces an effect similar to understeer intervention with a torque-vectoring actuator. This only requires a hydraulic modulator with a long service life and low noise generation instead of an additional torque-vectoring actuator.

Influencing self-steering properties through roll stabilization
Function
In line with their comfort-oriented basic function, systems for roll stabilization serve to compensate the rolling motion of the vehicle body when cornering.

Driving-dynamics applications
Where the roll-stabilization system has a two-channel design with separate activation of the actuators on the front and rear axles, there is an additional option for influencing the self-steering properties. The roll-compensating torques can – within the framework of the performance of the individual channels – be actively distributed between the front and rear axles. This results in a change of the wheel downforces.

The effect on the self-steering properties is based on the degressive increase in the lateral forces with the normal forces (Figure 3). When the rolling moment is extensively supported over the rear axle, the normal force at the rear on the wheel on the outside of the curve increases dramatically; the associated side force, however, increases only underproportionally. The side force on the rear axle is weakened, and the understeer tendency decreases. As a countermove, the understeer tendency increases when the rolling moment is extensively supported over the front axle.

Influencing self-steering properties through variable vibration dampers
Function
The use of variable dampers for adaptive damping of the vertical movement of the vehicle body has become significantly widespread. Pitching and rolling motions which are caused by driver inputs such as braking and steering or by road-surface defects, as well as vertical movements caused by road-surface defects, can be noticeably reduced.

Driving-dynamics applications
It is also possible with limited effect to influence the self-steering properties. The action mechanism corresponds to influencing the self-steering properties through roll stabilization. However, the function only has a transient effect during a rolling motion of the vehicle body, since a damper movement is a prerequisite for damping forces. The damping forces can be modulated by adjusting the damping hardness so that the wheel downforces can be semiactively distributed as described above.

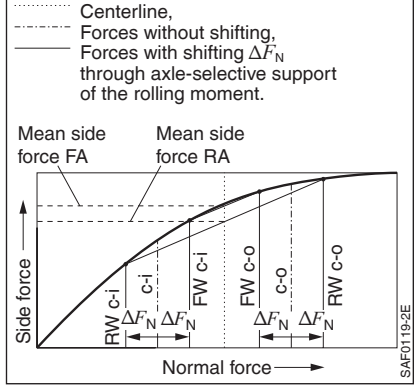
System architecture

Allocation of functions to ECUs
The actuators in the chassis and in the drivetrain are generally activated not only by functions of the integrated driving-dynamics control systems, but also first and foremost by basic functions which are very closely connected with the actuator. Examples of such basic functions include servo assist in an electrical power-steering system or variable steering ratio in an override steering system.

The basic functions are characterized by moderate networking with other vehicle systems. Because of their close connection to the actuator, they are usually integrated into the associated ECU, which also activates and monitors the actuator.

On the other hand, the functions of the integrated drivingdynamics control systems are highly networked, in particular with driving-dynamics control (ESC) and other chassis control systems. They calculate a setpoint driving-dynamic value for the actuator, which is transmitted via a data bus to the ECU. An arbitration with the setpoint values of the basic functions is performed in the ECU.

Figure 3: Side force as a function of normal force with a constant slip angle
Reduction of understeer tendency by shifting of normal wheel forces.
FA Front axle, RA Rear axle,
FW Front wheel, RW Rear wheel,
c-i Inside of curve, c-o Outside of curve.



The ECU for driving-dynamics control (ESC) or a central ECU of the "Chassis" functional area is suitable for use as the integration platform for the functions of the integrated driving-dynamics control systems (Figure 4).

Interaction of various functions

Various functions and actuators are increasingly being installed in a single vehicle. This calls into question a function structure which guarantees good interaction of the individual functions and which above all prevents mutual interference.

The original trend observed of implementing a separate controller for each actuator is reaching its limits. However, fully bringing together all the algorithms in a central controller limits flexibility in the distributed development and in the use of ECU resources.

A more promising compromise is bringing together all the controllers which utilize a related positioning principle, since interventions with a related positioning principle require particularly intensive coordination.

The following are mentioned as positioning principles and allocated positioning systems:

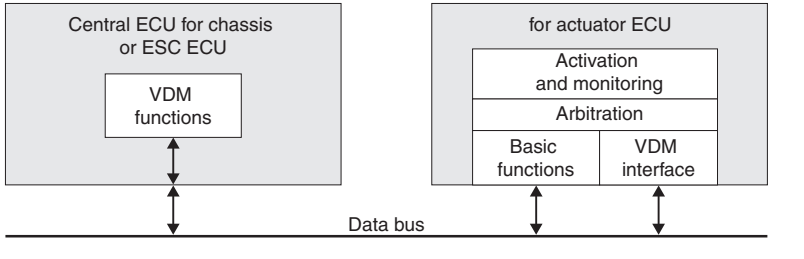
– wheel torque: driving-dynamics control, controllable differential locks, torque vectoring,
– steering angle: override steering (front axle) and rear-axle steering,
– normal force: roll stabilization, adjustable shock absorbers.

The driving-dynamic steering-torque recommendation plays a special role. It must be matched in terms of its effect on the driver with functions for override steering.

References

[1] Isermann (Editor): Fahrdynamikregelung – Modellbildung, Fahrerassistenzsysteme, Mechatronik. Vieweg Verlag, Wiesbaden, 2006.
[2] Deiss, Knoop, Krimmel, Liebemann, Schröder: Zusammenwirken aktiver Fahrwerk und Triebstrangsysteme zur Verbesserung der Fahrdynamik. 15th Aachen Colloquium Automotive and Engine Technology, Aachen 2006, PP. 1671…1682.
[3] Klier, Kieren, Schröder: Integrated Safety Concept and Design of a Vehicle Dynamics Management System. SAE Paper 07AC-55, 2007.
[4] Erban, Knoop, Flehmig: Dynamic Wheel Torque Control – DWT. Agility enhancement by networking of ESC and torque vectoring. 8th European All-wheel Congress, Graz 2007, PP. 17.1…17.10.
[5] Flehmig, Hauler, Knoop, Münkel: Improvement of Vehicle Dynamics by Networking of ESC with Active Steering and Torque Vectoring. 8th Stuttgart International Symposium Automotive and Engine Technology, Report Vol. No. 2, PP. 277…291, Vieweg Verlag, Wiesbaden 2008.

Figure 4: System architecture for implementing functions of integrated driving-dynamics control (VDM)



Vehicle bodies, passenger cars

Introduction

The vehicle body or the body shape determines the external shape of a vehicle and is itself essentially determined by the vehicle purpose or the use and the addressed customer group, but also by social, technological, and design trends. Design trends and the combination of body-shape archetypes have consistently given rise to new body shapes. One example of this is the Sports Utility Vehicle (SUV), which developed in the USA from the “station wagon” (estate car) and the classic off-road vehicle. Even new drive concepts, such as for example the electric drive, influence the realized body shapes.

Vehicle body and vehicle concept, described by the essential main dimensions and the positioning of the main components and the occupants, are closely interlinked. For this reason, first the body shapes and then the development of the vehicle concept or the vehicle architecture will be discussed in the following. To conclude, the effects of current technological trends such as for example automated driving and virtual product development on vehicle conception and vehicle bodies will be described.

Milestones in automobile history

The history of the automobile (or passenger car) has been marked by many individual discoveries, which will only be briefly described here. New product and process technologies have consistently led to innovations and trend reversals in the development of the automobile.

In 1886 Carl Benz and Gottlieb Daimler, independently of each other, built the first motor carriages with an internal-combustion engine and thereby initiated the triumphal advance of motorized private transport. However, by the end of the 19th century it had still not been decided which drive would gain acceptance in the automotive industry.

Already in 1834 Thomas Davenport in the USA built the first electrical DC motor and demonstrated its operation using an electrically powered model locomotive which was equipped with a non-rechargeable battery. In 1837 Davenport in the USA was awarded a patent for his electric motor. At the same time and independently in 1834 Moritz Hermann von Jacobi also developed an electric motor with a power output of approximately 15 W. In 1835 the two Dutchmen Sibrandus Stratingh and Christopher Becker constructed an electric motor which was used to drive a small model vehicle – the first demonstrable use of an electric motor. In 1881 the first electrical three-wheeler was demonstrated by Gustave Trouvé at the International Exposition of Electricity in Paris, i.e., already five years before the first automobile with internal-combustion engine. In 1882 Werner Siemens ran on a test track in Berlin the “Electromote”, the forerunner of today's trolley buses. In 1888 Andreas Flocken developed the first electric car from Germany with pneumatic tires and electric headlamps and a top speed of 15 km/h. In 1897 electric taxis were introduced in New York and by 1899 already 90 % of all the taxis in that city were electrically powered.

Up to the first decade of the 20th century automobiles with steam drives were

also built as alternatives to the internal-combustion engine drive and the electric drive. Examples include the White Touring Car from the White Motor Corporation and the Stanley Steam Car from the Stanley Motor Carriage Company.

Only with the invention of the electric starter in 1909 by Kettering in the USA and through further development in 1914 by Bosch in Germany (patent in 1914 to Bosch) and due to the rapid expansion of a filling-station network for mineral-oil-based fuels was the internal-combustion engine able to gain acceptance as the dominant drive for automobiles and push electric and steam vehicles into niche markets.

The invention of the pneumatic tire in 1888 by Dunlop in the UK and the invention of the steering wheel in 1900 by Packard in the USA are further milestones on the road to today's automobile. In 1914 Ford introduced assembly-line production with the Model T. As a result, automobiles could be produced in large numbers and were able to conquer the roads [1].

Not only the drive technology but also the configuration of the drive has influenced the automobile body shapes. The rear-wheel drive is the oldest form of drive and refers to a vehicle with a powered rear axle irrespective of the layout of the engine. In the standard drive the engine is housed in the front end and drives the rear axle through a propeller shaft with cardan universal joints via the differential. This makes it necessary to have a transmission or center tunnel and gives rise to weight and installation-space disadvantages. On the other hand, the standard drive facilitates the separation and splitting of the steering function and the drive function to one axle each and thereby results in better weight distribution. The rear axle is also powered in the rear-wheel drive. Here the rear engine, transmission, and drive axle form a complete drive unit at the rear of the vehicle. A front-wheel drive in conjunction with a front engine is generally referred to as a front-wheel drive. Further drive configurations are the all-wheel drive with two powered axles, the rear-wheel drive with mid engine, and the rear-wheel drive with transaxle configuration. The front-wheel drive and the standard drive have predominantly gained acceptance today.

Since the 1990s the electrification of the drive has experienced a renaissance. However, this was initially confined to concept vehicles and pilot tests, e.g., on the island of Rügen (1992 – 1996). With the foundation of the National Platform for Electric Mobility in Germany in 2010 electromobility based on lithium-ion-battery technology gathered significant pace again. In 2012 Tesla with its Model S launched the first luxury sedan with electric drive and frunk (front trunk, i.e., luggage compartment at the front of the vehicle) onto the market. Today electrification of the drive and the trend towards automated vehicles are opening up new possibilities for fashioning vehicle body shapes.

In more recent developments body shapes have become increasingly diversified. Each era has had its preferred body shapes. Thus in the 1980s Chrysler with the Chrysler Voyager popularized the family van or people carrier (Chrysler Minivan 1985). In the 1990s through to the first decade of the 21st century the Minivan experienced a revival as an automobile body shape with optimized space utilization and a high roof line. Mercedes-Benz with the A-Class brought onto the market a family sedan in the compact segment with an allowance for alternative drives in the vehicle floor with the look of a Minivan (one-box design). Towards the mid-1990s the Sports Utility Vehicle (SUV) arrived on the German market in the form of the Toyota RAV 4. Starting with large vehicles, the SUV in the 2000s conquered further segments extending down to compact vehicles [2]. In 2021, according to the Federal Motor Transport Authority (KBA) in Germany, SUVs held a market share of around 25 % of passenger-car sales.

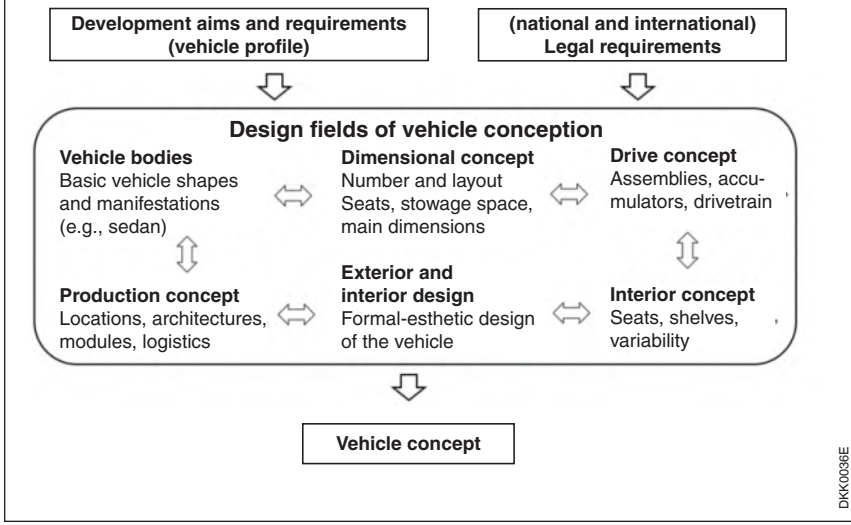
Definitions

Vehicle architecture
The vehicle architecture constitutes the totality of the components and systems of a vehicle family which facilitates a derivation of all the planned derivatives and body shapes of that vehicle family. Legal requirements, market and customer requirements, functional requirements, ergonomic requirements, safety requirements, production requirements, and economic requirements are all taken into account in the definition of vehicle architecture. The focus in vehicle architecture is on the concept-determining modules, components, and systems.

Vehicle concept
"The vehicle concept the structural design of a product idea with which the fundamental feasibility is safeguarded" [3]. The vehicle concept describes the vehicle in its essential main dimensions (inside and outside) and the positioning of the main components and the occupants (driver and passengers) in the vehicle. Over the years different body shapes have developed in response to the arrangement of the drive components. The vehicle concept ensures the technical feasibility of the arrangement of the assemblies and components and compliance with statutory provisions. This relates to the regulations covering crashes, e.g., deformation travel, and covering ergonomics, e.g., the accessibility of operating and control elements and outward visibility [4]. The design fields of the vehicle conception integrate various subconcepts, such as for example the drive concept, the interior concept, the production concept, and the formal-esthetic design of the vehicle (see Figure 1). The emphasis here is on the definition and design of body manifestation, basic vehicle shape, and future variants [3].

Vehicle body
The vehicle body defines the external shape of a vehicle and is established by the German standards DIN 70010 (System of road vehicles) and DIN 70011 (Bodies for passenger cars; terms and definitions) and the international standard ISO 3833 (Road vehicles – Types – Terms and Definitions). The standards differ from each other in the details. The vehicle body

Figure 1: Design fields of vehicle conception



or the body shape is determined by the vehicle purpose and the addressed customer group.

Vehicle package
"The package is the gradually refined elaboration of the design during the development of the vehicle with the aim of constantly check the technical feasibility of the planned product and the dimensional interaction of all the assemblies and components" [3]. This entails as a matter of priority the optimal use of the installation spaces in the vehicle. The different conflicts of aims, space requirements, and functional dependencies are weighed up and carried over into an arrangement that satisfies all the components geometrically and physically. This arrangement of all the components in the vehicle is called the vehicle package or package.

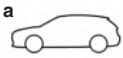
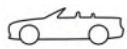
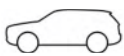
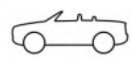
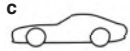
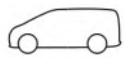
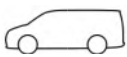
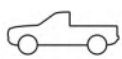
Vehicle body and vehicle segments

The vehicle body or the body shape is primarily determined by the vehicle purpose and the addressed customer group. Styling trends, social trends, and statutory provisions as well have consistently given rise to new body shapes. Historically, new vehicle body shapes have frequently come about as the result of elements from existing body shapes being put together in new ways. This is shown particularly in the so-called "cross-over body shapes", such as for example the Sports Utility Cabriolet (SUC), which combines a Sports Utility Vehicle (SUV) with a cabriolet or convertible. Elements from SUVs can be found today in many hybrid body shapes. The German Federal Motor Transport Authority (KBA) has defined vehicle segments for the carriage of persons and goods [5], which are already featured elsewhere (see Complete motor-vehicle system). This classification has been used to explain the vehicle body shapes and vehicle segments in Figures 2a, 2b, and 2c. Also customary internationally are still further classifications of vehicles into segments which follow a similar logic, e.g., the vehicle segments of the EU Commission or the vehicle-size categories of the NHTSA (National Highway Traffic Safety Administration) in the USA and other.

Figure 2a shows how the classic body shape of the sedan, which was previously the dominant body shape, has continued to diversify starting out from the three basic shapes of notchback, squareback, and station wagon/estate.

The development of the SUV body shape from the classic off-road vehicle with elements of sedan-like body shapes is an example of the creation of new body shapes. SUVs are characterized by a high seat position, increased ground clearance, an at least disengageable all-wheel drive, and a characteristic body shape based on an off-road vehicle. Over time SUVs have been able to gain acceptance as an alternative to the high-capacity sedan (Family Van) and establish itself as a new body shape across all vehicle-size categories [2]. This development continues and results, for example by borrow-

Figure 2: Vehicle body shapes and vehicle segments
a) Vehicle body shapes and vehicle segments using a sedan as the example,
b) Development of body shapes and vehicle segments on the example of an SUV,
c) Further body shapes and vehicle segments.
× common
(×) rare

Body shapes and manifestations	Description of the body shapes	Minicar	Subcompact	Compact class	Mid-range	Upper mid-range	Luxury class
a 	**Squareback** Body shape, 2-box design, with tailgate for simple payload	×	×	×	(×)		
	**Notchback** "Classic" sedan body shape			(×)	×	×	×
	**Coupé** 2-door sedan with flowing-flat roof line			(×)	×	×	×
	**Fastback** 4-door, coupé-like roof shape usually with roof lid			(×)	×	×	
	**Station wagon/estate** Body shape of a passenger car with large load-compartment space			(×)	×	×	
	**Open** Roadster (2-seater) or convertible (usually 4-seater) – derived from notchback or squareback vehicles	(×)	×	×	×	×	×
b 	**Off-road vehicle** Robust vehicles, plenty of ground clearance and large overhang angle		(×)	×	×	×	(×)
	**SUV** Partly off-road vehicles, more ground clearance and overhang angle		(×)	×	×	×	×
	**SUV coupé** Partly off-road with coupé-like flat silhouette		(×)	×	×	×	×
	**SUV cabriolet (SUC)** Cross-over, derived from SUV and convertible				(×)	(×)	(×)
c 	**Sports car** 2-seater passenger car with relatively low overall height and high performance			(×)	×	×	×
	**Supercar** 2-seater passenger car with very low height, optimized for max. performance						×
	**MPV** High-capacity sedan with more than 5 seats and large available space			×	×	×	
	**Van** Minibus with 5 to 9 seats and large, variable available space				×	×	(×)
 DKK0037-2Y	**Pickup** Cross-over derived from SUV and light-duty truck, widespread in the USA			(×)	×	×	×

ing from other classic body shapes like the Coupé, in the new body shape of the Sports Utility Coupé. An example of an SUV Coupé is the Mercedes-Benz GLE. There is a trend, particularly in electric vehicles, towards SUV Coupés. Examples of SUV Cabriolets (SUCs) are the Range Rover Evoque Convertible and the VW T-Roc Convertible.

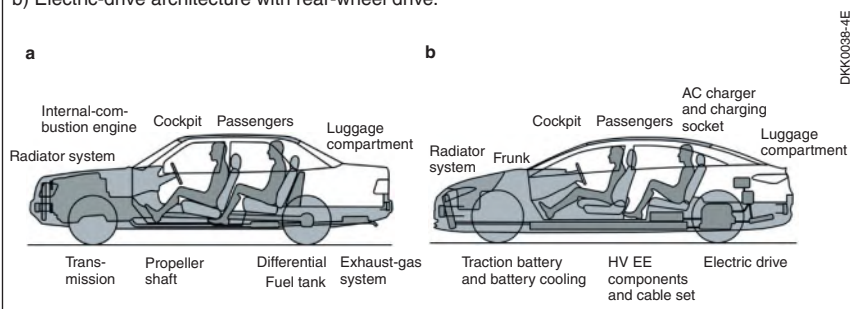
There are other relevant body shapes which are depicted in Figure 2c. These are, firstly, the emotive body shapes sports car and supercar and, secondly, the practical body shapes like the Multi-Purpose Vehicle (MPV) and the high-capacity sedan, the high-capacity van (minibus), and the pickup as a cross-over from the SUV and the light-duty truck.

Whereas the high-capacity sedan, e.g., the Renault Espace, is a family car with plenty of room and a variable-folding seating concept which is derived from a passenger-car architecture, the high-capacity van is a high vehicle derived from a light-utility-van architecture which is often characterized by removable seats mounted on continuous rails. The VW Bulli minibus was definitive of the high-capacity van. Today the segment is widely diversified.

Vehicle systems

Each vehicle package comprises the same system groups. These are the groups occupants, interior and stowage space, drivetrain, body structure, suspension and chassis, and wheels and tires (Figure 3). Each of these system groups can take on different manifestations depending on the functional objectives of the vehicle. Each of the vehicle systems has, through further development of the requirements, the potential to generate new body shapes.

Figure 3: Main components and systems of a passenger car (examples)
The drive determines the architecture and the dimensional concept.
a) Internal-combustion-engine architecture with standard drive,
b) Electric-drive architecture with rear-wheel drive.



Interface management in complete-vehicle development

The complete-vehicle conception and thus also the design of the vehicle body are characterized by the interdisciplinary cooperation of a number of technical disciplines and their interfaces. This is shown among others in the diverse requirements imposed on the vehicle concept and the package (Table 1) [3].

The first disciplines involved in the vehicle concept are complete-vehicle design, vehicle design (styling), and vehicle aerodynamics. But also further functions like drive, body design (body structure), and vehicle safety determine the vehicle concept. Here complete-vehicle design has an integrating function along the vehicle-conception process.

From a design viewpoint, the vehicle design can be divided into three types:
- the one-box design (e.g., a box-body light-utility truck or an MPV, e.g., Renault Espace),
- the two-box design (e.g., the square-back sedan, hatchback, e.g., VW Golf, station wagon or SUV),
- and the (classic) three-box design of the notchback sedan (e.g., Mercedes-Benz E-Class) [3].

The vehicle-design type constitutes the basis for all the design and constructive considerations with regard to styling, surface, and the like [6]. But also the drivetrain and the drive assembly with its peripheral components (e.g., accessories) determine via the vehicle package its interior and exterior design [6].

The complete-vehicle conception with its main functions of dimensional conception, ergonomics, and packaging embodies the "technical knowledge" in vehicle-body development. It is responsible for complying with the statutory provisions and integrating the requirements from design, aerodynamics, vehicle safety among others.

The vehicle aerodynamics and the design fashion the shape of the vehicle iteratively with different targets. The vehicle design process focuses on the design of the exterior and the interior. Here holistic considerations relating to the product environment, the customers, their use cases, and general design trends leave their mark. The vehicle design process is shown in Figure 4. Premises and an initial profile are elaborated on the basis of an initial analysis. Potential customers are described in the form of "personas" (fictitious customers representative of a particular target group). This is followed by an initial idea generation (brainstorming). The ideas are visualized in the form of sketches and a first technical dimensional concept and package are created in 3D. This is followed by the concept phase, in which the 3D designs are further elaborated and in part converted into hardware (1:4 clay models, 3D models). Methods of digital shape verification (VR, Powerwall) are used in the concept on a supporting basis. In the detailing phase a 1:1 clay model and a seat box are created in parallel to the design of the digital 3D model. The shapes found are returned by model scanning to the digital world. A model verification is performed in parallel

Table 1: Vehicle-concept and package requirements

	Requirements
Vehicle concept and package	Dimensional concept: – External dimensions – Interior dimensions
	Portfolio positioning
	Design, styling
	Aerodynamics and aeroacoustics
	Drive concept
	Chassis concept and wheels with tires
	Interior concept including operating and display concept
	Electrics/electronics architecture
	Safety concept (active and passive safety)
	Assembly, maintenance, and repair concept
	Sustainability including "end of life" concept
	Economic efficiency: – Use of funds (investments) – Product costs – Total cost of ownership (TCO)

by comparison with the technical "hard points" (e.g., dimensions of technical assemblies or lines of sight prescribed by law). This is followed by the so-called surfacing (Strak) phase (construction of class-A free-form surfaces), in which the surfaces are optimized from an optical and design perspective.

The most important target variable of vehicle aerodynamics is aerodynamic drag, supplemented by the output at the rear end, the aeroacoustics (responsible for wind noises), and the avoidance of contamination. From the aerodynamics perspective vehicles are blunt bodies which move with minimal ground clearance above the road surface [7]. The flow around the vehicle is three-dimensional and characterized by turbulent boundary layers. Flow separations must be avoided as far as possible because they result in increased flow resistance. Behind the vehicle a dead-water zone is created which is trailed by a wake. In the case of passenger cars powerful longitudinal vortices may also be created which interfere with the dead-water zone and the wake and contribute substantially to the flow resistance [7]. To reduce the energy demand of a vehicle and with it the fuel or energy consumption the vehicle's body shape is aerodynamically optimized in the direction of low c_d values (drag coefficients) and thus retroacts on the design of the vehicle.

The other drag-determining variable, the cross-sectional area of the vehicle, is determined by the body type (see Vehicle segments) and the number and posture of the occupants. The determination of openings for the cooling-air flows to deliver efficient thermal management for the drive and for further assemblies and components and to ventilate the passenger compartment constitutes further designs fields of vehicle aerodynamics [7].

Process of vehicle-concept development

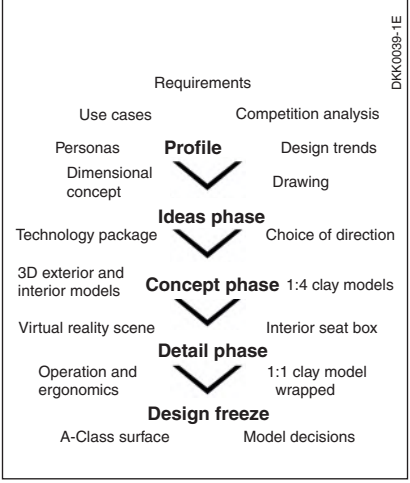
Vehicle concept

From the perspective of vehicle development the vehicle concept is at the start. It has the greatest influence on the vehicle body shape. Vehicle body shape and vehicle package, i.e., the accommodation of the assemblies and components in the vehicle and the interior design for the driver and the passengers are closely interlinked [3]. The vehicle concept comprises the following subconcepts ([3], [8], [1], [6]):

- drive concept and assembly position (front-, rear- or all-wheel drive, internal-combustion engine, hybrid or electric drive, front- or transversal installation of the drive assembly, etc.),
- interior concept (number of seats, comfort features, e.g., legroom, headroom, and stowage-space volumes),
- chassis concept (steered and rigid axles, suspension, wheels, and tires),
- thermal management and air-conditioning concept (heat recovery and heat utilization, active air conditioning, heat pump, thermal accumulators among others).

The drive concept and the chassis concept form the chassis and can be characterized as a drive platform which can be

Figure 4: Vehicle design process



used across a number of body variants. Volume manufacturers today develop body variants for model families which use common concepts and components. These are called the vehicle architecture. The interior concept and the drive concept have the greatest interactions with the sought-after body shape.

On the basis of the general vehicle concept – consisting of the subconcepts drive concept, chassis concept, body concept among others – the main dimensions of the vehicle such as vehicle length, vehicle width, and vehicle height as well as the interior dimensions are defined [9]. Selected external dimensions as per SAE or ECE are shown in Figure 5.

In the interior an empirical relationship between interior height and interior length can be demonstrated. In taller vehicles the occupants are seated in a more upright position and consequently the vehicle interior can theoretically have a shorter design [9]. Limits to this correlation are presented by the legroom and kneeroom of the occupants. Selected internal dimensions are shown in Figure 6.

The path from the first strategic considerations to the finished vehicle concept covers a number of steps. The first vehicle package should be kept simple and only show the essential concept features. With regard to the vehicle body, the first so-called "hard points" are defined. These are definitions of the main dimensions which are set for legal, strategic, design or technical reasons. The vehicle proportions are defined by few boundary conditions which are set in the vehicle profile, such as for example the number of occupants, the drive type and drive position, the wheel and tire sizes and the size of the luggage compartment, the ground clearance, and the sought-after crash-safety level. These definitions are referred to in the following as functional vehicle objectives [1].

Conception steps

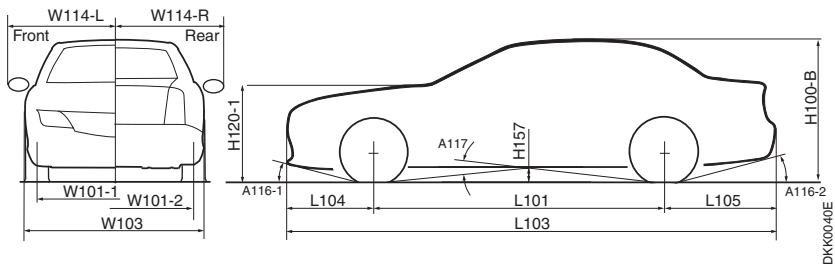
Step 1

The design process starts with the idea generation for design and package (see Interface management and Complete-vehicle development) and is geared towards the functional objectives for the vehicle concept. Even initial considerations regarding the body structure and the doors, gates, and lids have some bearing on the initial design. Further elaboration of the vehicle body in conducted from the inside outwards.

Step 2

In the second step the occupants of the first seat row (driver, front passenger) are positioned in the vehicle. The height-dimension chain is developed from the ground and takes into account the sought-after ground clearance H157 (closely correlated with the selected vehicle segment), the package of assemblies in the vehicle floor (e.g., energy storage devices), and the segment-specific sitting posture of the occupants in the vehicle. The sitting posture is determined by the seat reference point (R point) or the associated H30 dimension, the heel point,

Figure 5: Selected external dimensions as per SAE or ECE
(Front/rear view and side view.)



and the position of the steering wheel. With regard to the driver's sitting posture, statutory provisions must be observed (Figure 7), such as for example forward visibility, but also functional and ergonomic considerations have a bearing on the choice of seat position, such as for example the aerodynamics of the resulting vehicle body and the ergonomics of the entry/exit. Digital human models representing specific percentiles of the height distribution of potential customers are used in the validation of the requirements [10]. It has proven worthwhile for layout purposes to validate the requirements with a tall 95th percentile man (95 % of men are shorter or of equal height) and a short 5th percentile woman (fewer than 5 % of women are shorter) ([4], [1]).

Step 3

In the third step the rear passengers are positioned. Here today – depending on the vehicle segment – one or two rear-seat rows are conceivable. The number of passengers in the rear who sit next to each other determine the vehicle width at the height of the respective seat row. For example, two individual seats, two full seats and one emergency seat or three full seats are conceivable in the rear. Leg-room, headroom, and shoulder room of the rear passengers are verified with digital human models.

Entry/exit simulations or trials with test subjects can be used to validate and optimize the first conceptional definitions in the course of the vehicle-concept development.

Step 4

In the fourth step the final decision regarding the drivetrain is made and this is located in the vehicle layout. Depending on the drive type, this concerns the components driving machine (internal-combustion engine, e-machine with inverter), energy storage device (fuel tank, traction battery), and miscellaneous components of the drivetrain (e.g., drive shafts) [1]. The choice of drive type, the number of powered axles, and the position of the drive have concept-determining effects on the vehicle proportions and are thus characteristic of specific vehicle segments. New

Figure 7: Layout of driver's cockpit
Passenger-car steering-wheel position as per DIN 70020 (Automotive engineering – General terms; definition and explanation).
R point: Seat reference point,
AHP: Accelerator heel point.

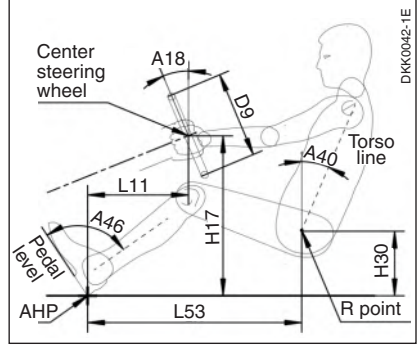
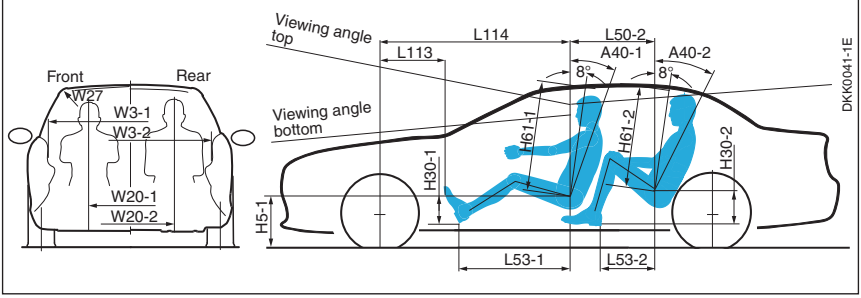


Figure 6: Selected internal dimensions as per SAE or ECE
(Front/rear view and side view.)



drive types or drive configurations result if necessary in new vehicle body shapes.

Step 5
In the fifth step the position of the occupants is established in the y-direction (vehicle width), having regard to the packaging of the drivetrain, the crash behavior, and the comfort requirements. This definition is crucial to the feeling of space and comfort in the interior, but also to the safety of the passengers in the event of a side collision. In this phase a rough package of the occupants and all the essential assemblies should verify the rough dimensional concept in the essential dimensions [1]. In practice the individual steps can also take place on account of project-specific settings in a different order. Statutory provisions with regard to width and height of the vehicle body must be verified and observed.

Step 6
In the sixth step the installation spaces are defined for the payload. Depending on the requirements profile the length of the overhangs is influenced by the required load-compartment space and in turn gives rise to segment-specific body shapes, e.g., station wagon. New drive concepts here open up new stowage-space options, such as for example the so-called frunk (front trunk). In the development of vehicle architectures different variants are conceivable in a vehicle model. The position of the powered axle(s) in relation to the position of the main drive has concept-relevant consequences, e.g., the need for a propeller shaft to direct the power flow from the drive to the powered axle. This requires installation space in the interior (so-called tunnel) and in the underbody area as well as appropriate ground clearance.

Steps 7 and 8
In the next steps 7 and 8 the wheelbase and the wheel and tire dimensions are finally established. The distance between the vehicle axles determines the weight distribution and the package efficiency as well as the comfort of the occupants. Here the benefit dimensions of the vehicle and also the sought-after price positioning (volume market vs. premium market) must be observed. Low-priced subcompact-size cars are generally built as compactly as possible. This produces a short vehicle length, but also a relatively confined vehicle interior. For premium vehicles the wheelbase is chosen in such a way as to create a generous feeling of space. Particularly in electric vehicles with long ranges the battery installed in the underbody is decisive for the wheelbase. Boundary conditions such as the turning circle also have a bearing on the dimensioning of the wheelbase. The ground-clearance requirements are also verified in this step.

Step 9
In the ninth step the respective track of the front and rear axles is established. Particularly in the rear area there are here conflicts of aims between the space requirement of the passengers and the axles, wheels and tires on the one hand and the aerodynamics, which can be improved by reducing the vehicle width at the rear, and the stylistic body design of the vehicle.

Step 10
In the tenth and last conception step the vehicle (cross-)sections are established at the height of the axles, at the height of the driver and front passenger, at the height of the rear passengers, and if necessary in the area of the luggage compartment [1]. In an iterative harmonization process between technology and design the optimal shaping of the vehicle exterior is established in the field of conflict of different requirements. The exterior surfaces are influenced by the body structure, the arrangement and function of the doors, gates, and lids (door concept), and the interior layout (interior concept).

Sectional elaboration of the vehicle concept

Installation-space sections
The sections through the vehicle divide the vehicle installation space into sections which can be elaborated with segment- and body-specific targets (Figure 8) [1]. Alternatively, digital 3D package rough models are used for the first design.

Section 1
The first section up to the vehicle bulkhead (after the front axle) is characterized for example for a front-wheel drive by the package of the drive assembly and the design of the crash structures. The position of the drive assembly (longitudinal or transversal) and the crash structure necessary to satisfy the front-collision requirements are the determining issues in this installation-space section.

Section 2
The front passenger compartment is the next installation-space section, which is essentially characterized by the needs and requirements of the driver's cockpit and the driving posture of the driver and the front passenger [6]. Here statutory provisions (regarding visibility) and the accessibility of the operating and control elements are concept-determining. An upright driving posture results in a high vehicle architecture and enables the wheelbase to be shortened ([1], [6]).

Section 3
The next section at the height of the B-pillar to the end of the vehicle interior de-

Figure 8: Concept-relevant sections through the vehicle

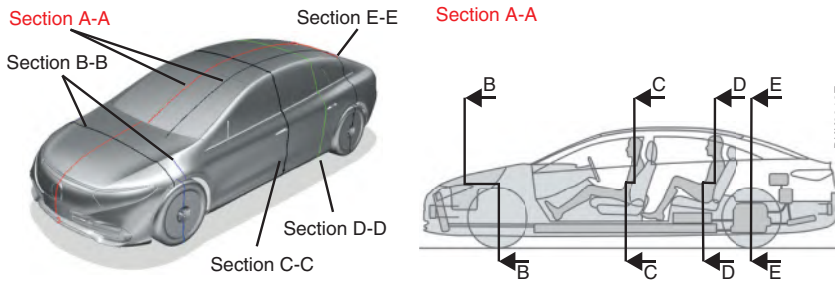
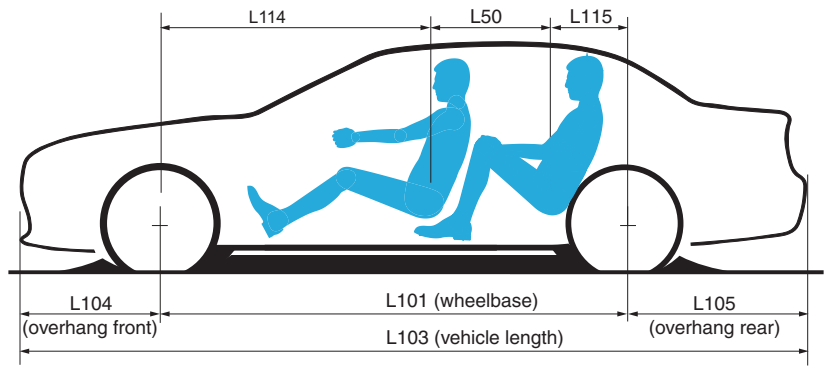


Figure 9: Dimension chain in longitudinal direction to define the vehicle length



scribes the rear passenger compartment. The number of rear seat rows, the sitting posture of the rear passengers, and their legroom and kneeroom determine the length of the wheelbase and thus also the vehicle length.

Section 4
The luggage compartment is the fourth and last installation-space section. It is determined by the required load-compartment space, the packaging of a fuel tank or a traction battery or if necessary a spare wheel, and the required crash length in the event of a rear impact.

Vehicle length
The four installation-space sections result in the vehicle length, which for its part is subject to segment-specific and legal restrictions. The dimension chain in the x-direction is shown in Figure 9.

Vehicles should adhere to the customary garage dimensions and length restrictions, for example for carriage on motorail trains or ferries. A generous vehicle length furthermore entails additional weight, additional costs, and driving-performance limitations. Therefore the complete-vehicle conception aims to minimize the vehicle length by adhering to all the requirements.

Current trends of vehicle-concept development

New technological and social developments influence vehicle conception. This changes the process of vehicle conception through new methods and tools and can also result in new body shapes.

Electrification of the drivetrain
The transition to electromobility means that a traction battery of a size commensurate with the range requirements of the customers must be accommodated in the vehicle. The battery modules can be installed in a distributed arrangement in the vehicle or as a large battery pack, e.g., in the underbody. In fuel-cell vehicles hydrogen accumulators must be incorporated aside from a smaller high-voltage battery and the e-machine(s) together with inverter. Energy storage devices, such as for example high-voltage batteries and hydrogen tanks, are today usually accommodated in the underbody. If in future the volumetric energy densities of energy storage devices increase due to technological advances, the installation-space situation will ease and lower architectures will be easier to implement.

Automated driving
Automated driving is one of the trends for the 21st century. This carries with it new installation-space demands for powerful computers and sensors, e.g., lidar sensors, video cameras among others, which are required for highly automated driving. On the other hand, highly automated driving could open up the opportunity to deliver the vision of accident-free driving. Technologies like drive-by-wire (electronic accelerator and brake pedals) and steer-by-wire (electronic steering) play a significant role on the road to automated driving. The automation of all driving functions leads via driver-assistance systems, which support the driver, to autonomous driving, which relieves the person completely of the driving function and responsibility and thus in future will make possible driverless transport systems on public roads ([11], [12]).

Digitalization of vehicle development
Increasing digitalization in vehicle developing opens up to vehicle constructors and designers new opportunities to develop vehicles digitally from the outset. This ranges from modern tools for quickly sketching and designing vehicle surface models through data-driven development, which makes possible in connection with "machine learning" the development of intelligent vehicle systems. The digital transformation of the vehicle-development process makes possible the collaborative integration of representatives of different specialist departments and divisions, even beyond corporate boundaries and national borders [13].

Virtual reality (VR)
VR tools – including headsets for augmented and mixed reality – used to be classed with video games and films. Today they accelerate vehicle development and enable vehicle-concept teams around the world to completely refashion vehicles and mobility experiences. Form finding for new vehicle models and the presentation of innovative vehicle concepts can take place in a virtual future, be experienced immersively, and be discussed and further developed with decision makers. Furthermore, VR tools play an increasingly more important role in the functional verification of vehicle concepts. Here the tools are used in a so-called mixed reality, where physical structures (hardware) and VR elements complement one another, or even in a purely virtual environment.

In future algorithms and expert systems will support subproblems of vehicle development, such as for example the choice of wheels and tires. The incipient states are already being tested. The future vision would be a completely AI-supported (Artificial Intelligence) vehicle-design process.

References

[1] S. Macey, G. Wardle: H-Point: The Fundamentals of Car Design & Packaging. Design Studio Press, 2nd Edition, 2014.
[2] S. Buhren: Eine kurze Geschichte der SUV. Online: https://www.handwerksblatt.de/mobilitat/eine-kurze-geschichte-der-suv; as at 11/2017.
[3] A. Achleitner, Ch. Burgers, G. Döllner: Fahrzeugkonzept und Package. In: Pischinger & Seiffert (Editors), Vieweg Handbuch Kraftfahrzeugtechnik, Chapter 4.2. pp. 140–170, Springer Vieweg, ATZ/MTZ-Fachbuch, 8th Edition, 2016.
[4] H. Bubb, R. E. Grünen et al.: Anthropometrische Fahrzeuggestaltung. In: Bubb et al. (Editors), Automobilergonomie pp. 345–470, Springer Vieweg, 2015.
[5] KBA: Verzeichnis zur Systematisierung von Kraftfahrzeugen und ihren Anhängern. As at: August 2019.
[6] H. Seeger: Basiswissen Transportation Design. Springer Vieweg, 2014.
[7] W.-H. Hucho: Aerodynamik der stumpfen Körper. Physikalische Grundlagen und Anwendungen in der Praxis. Springer Vieweg, 2nd Edition, 2011.
[8] J. Weissinger, T. Breitling: Fahrzeugkonzeption in der frühen Entwicklungsphase. In: Pischinger & Seiffert (Editors), Vieweg Handbuch Kraftfahrzeugtechnik, pp. 1276–1282. Springer Vieweg, ATZ/MTZ-Fachbuch, 8th Edition, 2016.
[9] H. Pippert: Karosserietechnik – Konstruktion und Berechnung. Vogel, 3rd Edition, 1998.
[10] H. Bubb: Menschmodelle. In: Bubb et al. (Editors), Automobilergonomie, Chapter 5, pp. 221–258. Springer Vieweg, 2015.
[11] H. Winner, S. Hakuli, G. Wolf (Editors): Handbuch Fahrerassistenzsysteme: Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort. Springer Vieweg, 3rd Edition, 2015.
[12] M. Maurer, Ch. Gerdes et al. (Editors): Autonomes Fahren: technische, rechtliche und gesellschaftliche Aspekte. Springer Vieweg, 2015.
[13] U. Winkelhake: Die digitale Transformation der Automobilindustrie. Springer Vieweg, 2nd Edition, 2021.

Vehicle bodies, commercial vehicles

Classification of commercial vehicles

Commercial vehicles are used for the safe and efficient transportation of persons and freight. In this respect, the degree of economic efficiency is determined by the ratio of usable space to overall vehicle volume, and of payload to gross vehicle weight. Dimensions and weights are limited by legal regulations.

A wide variety of vehicle types meet the demands of local and long-distance transportation, as well as the demands encountered on building sites and in special applications (examples in Figure 1). Commercial vehicles can basically be subdivided into the following categories: light utility vans, medium- and heavy-duty trucks, buses, tractor units, construction and agricultural machines, and special vehicles (e.g., piste basher, airfield fire truck).

Due to the wide diversity of vehicle types, the process of calculating the dimensions of the body structures (unitized body, cab, chassis, etc.) takes on a major importance right from the earliest stages of design. Building on experience with comparable vehicles, benchmark designs (volume-sales units, worst-case configurations) are defined by simulation and calculation with the aid of gradually refined complete-vehicle models using FEM (finite-element method) or MBS (multiple-body simulation). In this way, the rigidity, operational integrity, and vibration, acoustic and crash characteristics, etc. of relevant body-structure variants can be obtained to a substantial degree by computational means, even before testing starts. Structural calculations also take account of the requirements of (international) statutory safety standards.

Figure 1: Overview of commercial vehicles
(Examples).
a) Light utility van, b) Truck,
c) Articulated road train,
d) High-capacity road train,
e) Semitrailer (Europe),
f) Semitrailer (NAFTA), g) Bus.

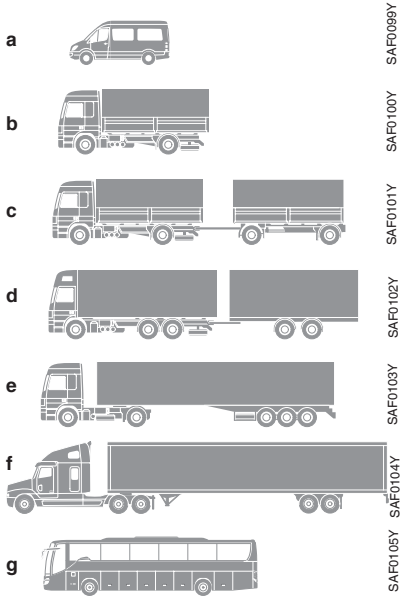
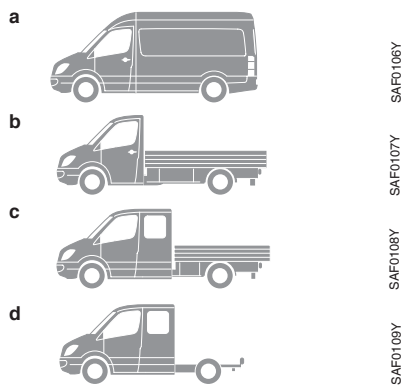


Figure 2: Overview of light utility vans
(Examples).
a) Panel van, b) Flatbed van,
c) Twin cab, d) Chassis.



Light utility vans

Areas of application
These are light utility vans (2 to 7 t) used in the transportation of persons and in local freight distribution. Light utility vans with more powerful engines are also increasingly deployed in pan-European long-distance transport duties involving high mileages (express delivery services, overnight courier services). In both cases, stringent demands are made on the vehicle in terms of agility, performance, user-friendliness, and safety.

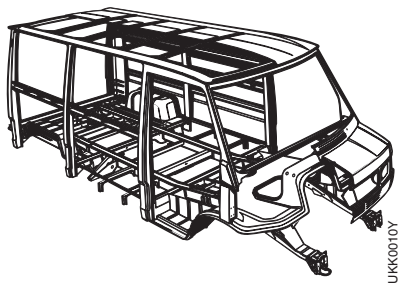
Body-structure variants of light utility vans
The design concepts are based on front-mounted engine, front or rear-wheel drive, independent suspension, or rigid axle and, with overall weights of over 3.5 to 4 t, twin tires on the rear axle.

The product range includes enclosed-body multipurpose panel vans and platform-body vehicles, as well as low-bed and high-bed platform trucks with special superstructures and crew cabs (examples in Figure 2).

Up to approximately 6 t overall weight, the bodies form an integral load-structure unit together with the chassis (Figure 3). The body and chassis frameworks consist of sheet-metal pressed elements and flanged profiles similar to passenger cars.

Light utility vans with platform bodies have a ladder-type frame with open or closed side members and cross-members as the primary load-bearing structure (similar to trucks, next section, Figure 5). These open designs also serve to build up box or recreational-vehicle bodies.

Figure 3: Light utility van, load-bearing unit



Medium- and heavy-duty trucks and tractor units

Body
Trucks are divided into four essential subtypes: long-distance transport vehicles, delivery vehicles, construction vehicles, and special vehicles. What all these vehicles have in common is a supporting frame structure on which a resiliently mounted cab is located and which constitutes a body which is connected with shear strength to the frame. The maximum dimensions and maximum permitted vehicle weights are subject to statutory provisions, which sometimes differ significantly from country to country. The maximum dimensions with regard to overall and semi-trailer lengths are stipulated for example in Europe by Directive 96/53/EU [1], the new version of which 2015/719/EU [2] is currently being technically elaborated. In the NAFTA zone, however, only the semi-trailer length, but not the overall length, is subject to restriction.

Equipped with a front-mounted engine, the drive is provided via one or more dual-tire rear axles. In individual cases, the rear axle is fitted with single tires. Four-wheel drive with an additional transfer case including interaxle-differential lock is also used for construction-site applications (off-road) with high traction requirements. The powered axles in this case have as standard axle-differential locks in their differentials and interaxle-differential locks for tandem-axle assemblies.

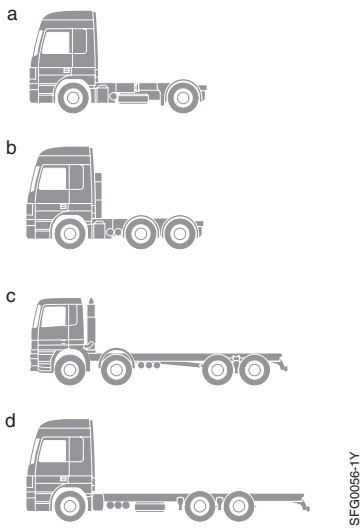
Chassis
Designations
The type of truck chassis (Figure 4) is given its designation in accordance with the convention N×Z/L, where N denotes the number of wheels or wheel pairs, Z the number of driven wheels or wheel pairs, and L the number of steered wheels (twin wheels count as one wheel pair). If L is not specified (e.g., 4×2), the vehicle has two steered front wheels.

Wheel suspension
Truck chassis have pneumatic- or leaf-spring-suspension rigid front and rear axles. Independent wheel suspensions are only used occasionally. Pneumatic

suspensions reduce the body acceleration to improve driving comfort, protect the load, and reduce strain on the road. They also facilitate the raising and lowering of swap bodies and the attachment and removal of semitrailers. These applications are usually used with 4×2 and 6×2 chassis for long-distance transport and delivery traffic.

Tandem-axle assemblies
For construction-site use or in the case of increased traction requirements the chassis are equipped with several powered axles (e.g., 4×4, 6×4, 6×6, 8×4, 8×8). In this case, the rear axles are combined to form a tandem-axle assembly. For the most part a steel-spring suspension is used, but pneumatic suspensions are also featured on the rear axles for road-based applications (e.g., building-material or low-loader transport) with the previously mentioned advantages. The axle load in a tandem-axle assembly is compensated in the case of the steel spring mechanically by pivoted mounting (center mounting) between the axles. In the case of non-combined pneumatically sprung single axles, the axle load is normally compensated pneumatically by altering the spring stiffness of the pneumatic suspension of the single axles.

Figure 4: Truck chassis
(Examples).
a) 4×2 (four wheels, two of which driven),
b) 6×2/4 (six wheels, two of which driven, four steered),
c) 8×6/4 (eight wheels, six of which driven, four steered),
d) 6×2 (six wheels, two of which driven).



Leading and trailing axles
Three-axle vehicles (6×2) are fitted with either a leading or a trailing axle (nonpowered axles in front of or behind the driven axle) to increase the payload capacity. If the chassis is fitted with a leading axle, this is mounted as an additional axle in front of the powered rear axle and provided with pneumatic suspension.

These leading axles usually have a steered design in order to avoid constraints when cornering, deliver a small turning circle, and thereby increase the maneuverability of the complete vehicle. The steering was effected in the first designs by a steering linkage branched from the steering axle, but is now equipped as a rule with electro-hydraulic steering because this does not require a linkage that takes up valuable space.

It is possible to equip the leading axle with a smaller wheel/tire size differing from the powered axle, but in this case the bearing loads of the leading axle are lower (4 to 5 t instead of a possible 7.5 to 9 t). There is also the combination of the leading axle with a lifting device. Then, when the vehicle is driving empty or with a low load, the axle is lifted from the driving position via a further air-spring bellows operating in the axially centered direction. This decreases rolling resistance and consequently reduces fuel consumption and also tire wear. Both these factors help to improve the vehicle's economic efficiency.

In the case of chassis designs with a trailing axle, this axle is mounted behind the powered rear axle. It is as a rule provided in on-road vehicles with pneumatic suspension and fitted with single tires, and in this case usually has a bearing load of 7.5 t. This variant can also be steered. Where trailing axles were still rigid when they began to be used in the late 1960s with the emergence of swap-body artic-

ulated road trains, steered trailing axles came to the fore with the desire for improved maneuverability against the backdrop of the increased use of three-axle chassis in delivery traffic. In view of the space conditions the trailing axle here is as a rule fitted with electro-hydraulic steering so that these variants can be flexibly used with different chassis and wheelbases. The use of a steered trailing axle offers, with a corresponding vehicle configuration, the combination of the load capacity of a three-axle chassis with the maneuverability of two-axle vehicle, which affords considerable benefits in economic terms. The trailing axles are also designed to be liftable virtually across the board with the previously mentioned advantages.

Where 6×2 chassis are operated on dirt roads or outside Europe, there are twin-tire rigid trailing axles with a bearing load of 9 to 10 t. As a rule, these are not designed to be liftable, but liftable variants are possible.

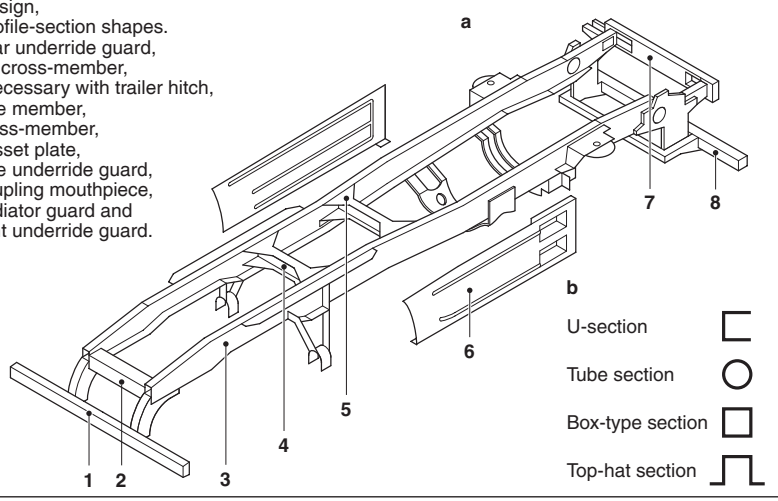
For special chassis the multiple fitment of leading and trailing axles or even the use of powered leading or trailing axles is conceivable. These chassis variants are conceived as special constructions and constitute technical challenges in that each axle must be assigned a different steering angle for the wheels so that the vehicle runs without excessive tire wear when cornering and complies with the turning-circle requirements. Combining a number of leading and trailing axles can thus also produce five- to eight-axle chassis which make possible gross vehicle weight ratings of over 60 t, for example for the mounting of concrete pumps or lifting work platforms with a working height of up to 112 m.

Chassis frame

The chassis frame is the commercial vehicle's actual load-bearing element. It is designed as a ladder-type frame, consisting of side members and cross-members (Figure 5). The vehicle body is bolted onto and the driver's cab is positioned on the chassis frame (Figure 6). The dimensions of the members are chosen to suit the required severity of application and load capacity (light- and heavy-duty commercial vehicles), but also in line with cost and weight considerations. The choice of profile sections (number and thickness) determines the level of torsional stiffness. Torsionally flexible frames are preferred in medium- and heavy-duty trucks because they allow the suspension to cope better with uneven terrain. Torsionally stiff frames are more suitable for light-duty delivery vehicles.

Figure 5: Ladder-type frame for a truck

a) Design,
b) Profile-section shapes.
1 Rear underride guard,
2 Tail cross-member,
if necessary with trailer hitch,
3 Side member,
4 Cross-member,
5 Gusset plate,
6 Side underride guard,
7 Coupling mouthpiece,
8 Radiator guard and
front underride guard.



Apart from the force introduction points, critical points in the chassis-frame design are the side- and cross-member junctions (Figure 7). Special gusset plates or pressed cross-member sections form a broad connection basis. The junctions are riveted, bolted, and welded. U- or L-shaped side-member inserts provide increased frame flexural strength and reinforcement at specific points.

The chassis frame also serves to accommodate by means of a variety of holders a wide range of add-on components such as fuel tanks, battery-device holders, compressed-air tanks, exhaust-gas system, or spare wheel. The configuration varies here, depending on the application profile required. Special mounted installations such as, for example, loading cranes or platform lifts are likewise mounted on the chassis frame for corresponding applications.

Figure 6: Truck assemblies
1 Cab, 2 Engine, 3 Transmission, 4 Axle, 5 Chassis frame, 6 Body.

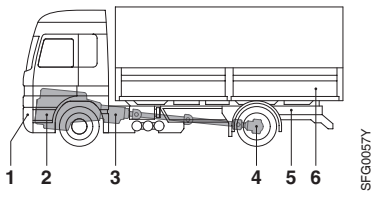
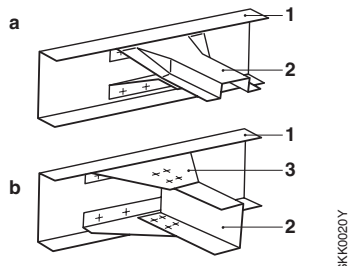


Figure 7: Junctions
a) Cap cross-member, b) U-cross-member.
1 Side member, 2 Cross-member, 3 Gusset plate.



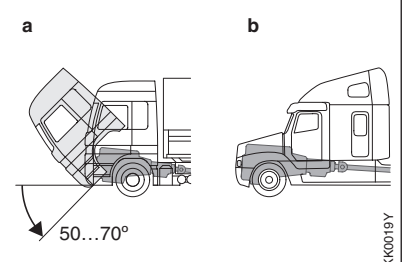
Driver's cab
There are a variety of cab designs available depending on the vehicle concept. In delivery vehicles and for public-service use, low, convenient entrances are an advantage, whereas in long-distance transport applications, space and comfort for example with a level floor are more important. Modular design concepts allow for short, medium and long cab versions while retaining the same front, rear and doors.

The cab is connected to the chassis frame by the cab mounting. A distinction is made here between comfort and standard mountings with different spring-and-damper combinations or transverse leaf-spring links with natural frequencies of 1 to 6 Hz.

From the design-concept viewpoint, a distinction must be made between cab-over-engine (COE) and cab-behind-engine (CBE) vehicles (Figure 8). In the case of cab-over-engine (COE) vehicles, the bulkhead and steering system are positioned right at the front of the vehicle. The engine is located under the cab (in the case of a level floor) or underneath an engine tunnel between the driver and the co-driver. The entrance is positioned in front of or above the front axle. A mechanical (by pretensioned torsion bars) or hydraulic cab-tipping mechanism provides access to the engine.

In cab-behind-engine (CBE) vehicles, the engine/transmission assembly is mounted ahead of the cab firewall under

Figure 8: Driver's cab
a) Cab over engine,
b) Cab behind engine.



a steel or plastic hood which is usually tiltable for reasons of accessibility. The driver enters the cab behind the front axle.

The standards required of the driver's cab with regard to aerodynamics, choice of materials, corrosion or equipment must be viewed as the same as the standards required of an automobile body. This means that, to save fuel and consequently CO_2 emissions, the aerodynamic drag of firstly the cab and secondly the entire truck-trailer must be reduced. Both the tractor vehicle and the trailer or the semitrailer must be included in the analyses for this purpose. Field tests show that including the semitrailer in the optimization of the complete-vehicle aerodynamics can achieve fuel savings of between 2 % and 4.5 %.

Body structures

Specific body structures such as flatbeds, box vans, dump-truck deep-beds, tankers, and concrete mixers permit the economical and efficient transportation of a wide variety of freight and materials. The body and load-bearing chassis frame are joined sometimes by means of auxiliary frames with non-positive or positive attachments. Special design features (e.g., sprung mountings in the forward body area) are required to connect the chassis frame (which usually has low torsional strength) to a rigid body (e.g., box-type). For off-road vehicles the torsional flexibility of the chassis frame must not be limited by a rigid body structure. Three-point mountings of the body structure are therefore used. For these vehicles coil instead of leaf springs are also used in the chassis to achieve large spring deflections.

Articulated road trains and semitrailer rigs are used in long-distance transport (Figure 1). As the size of the transportation unit increases, the costs relative to the freight volume decrease. Load volume is increased by reducing the empty spaces between the cab, body, and trailer (high-capacity road train, Figure 1d). Advantages of semitrailer rigs lie in the greater uninterrupted loading length of the cargo area and possible shorter inoperative times of the tractor units. Measures to improve aerodynamics, such as front and side panels on the vehicle and specially adapted air deflectors from the cab to the body, are applied to minimize fuel consumption.

The use of longer heavier vehicles which do not exceed the gross weight rating of 40 t but extend the overall road-train length to 25.25 m (instead of 16.50 m for semitrailers and 18.75 m for articulated road trains) can generate additional economic benefits. These longer heavier vehicles are used for light but bulky transported goods, e.g., for molded parts for the automotive industry, or in the transport of consumer goods.

Buses

The bus market offers a specific vehicle for practically every application. This results in a wide range of bus types, which differ in their overall dimensions (length, height, width) and appointments (depending on the application) (Figure 9).

Bus types

Microbuses

Minibuses carry up to approx. 20 passengers. These vehicles have been developed from light utility vans weighing up to approximately 4.5 t.

Mini- and midibuses

Buses which can carry up to approximately 25 persons are called mini- or midibuses. The transition from minibus to midibus is fluid. These vehicles have for the most part been developed from light utility vans weighing up to approximately 7.5 t. They are occasionally built on ladder-type chassis of light-duty trucks, or a unitized integral structure is used. A modified suspension design and special measures carried out on the body (e.g., flexible mountings) result in optimum ride comfort and low noise levels.

City buses

City buses are equipped with seating and standing rooms for scheduled routes. The short intervals between stops in suburban passenger transportation operations require rapid passenger turnover. This is achieved by wide doors that open and close swiftly, low boarding heights (approx. 320 mm) and low floor heights (approx. 370 mm).

Main specifications for a standard municipal-service bus:

– Vehicle length approx. 12 m,
– Permissible total weight 18.0 t,
– Number of seats 32 to 44,
– Total passenger capacity approx. 105 persons.

The use of double-decker buses (length 12 m carrying up to approximately 130 passengers), three-axle rigid buses (length up to 15 m carrying up to approximately 135 passengers), and three-axle articulated buses (length up to approximately 18 m carrying 160 to 190 passengers) provides increased transport capacity.

Now articulated buses with an overall length of almost 21 m and a total weight of 32 t are used in large numbers as city buses. These make use of an additional steered trailing axle which increases the bus' weight capacity and at the same time improves the maneuverability despite the greater overall length.

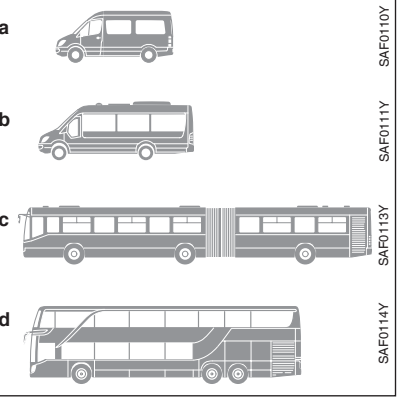
Intercity/overland buses

Depending on the application (standing passengers are not permitted at speeds over 60 km/h), low-floor designs featuring low boarding and floor heights (as in city buses) or higher floors and small luggage compartments (already very much like tour buses) are used. Intercity buses come in lengths of 11 to 15 m as rigid vehicles or in lengths of 18 m as articulated vehicles.

Tour buses (coaches)

Tour buses are designed to provide comfortable travel over medium and long distances. They range from the low, two-axle standard bus through to the double-decker luxury bus. Tour buses come in lengths of 10 to 15 m. Articulated buses were only occasionally used in intercity-bus applica-

Figure 9: Overview of buses
(Examples).
a) Microbus,
b) Mini- and midibus,
c) City bus,
d) Tour bus.



tions and did not gain acceptance. Double-deckers, on the other hand, are frequently used as intercity buses, particularly on long-distance bus routes which are operated with tour buses.

Trailer operation
Buses with trailers are increasingly being used. This applies to all applications, i.e., city and intercity/overland buses, as well as to tour buses.

In city-bus applications trailers are specially approved to carry passengers and at peak times are hitched to the towing vehicle to meet passenger demand. This type was used up until the 1960s, but disappeared very quickly with the emergency of articulated buses. Since around 2010 the city-bus/trailer combination has been back in favor so as to provide a flexible response to passenger fluctuations depending on the time of day on the outskirts of big cities.

With intercity/overland buses trailers serve to transport luggage or to carry bicycles in vacation regions so as to be able to offer an attractive and flexible means of transport for changing to local public transport.

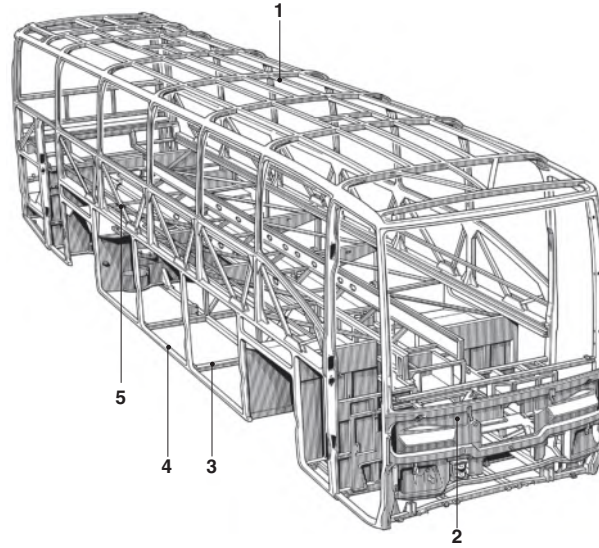
In the case of tour buses, particularly double-deckers, which in view of their design can offer minimal luggage space, the luggage trailer is a good solution as an economical and comfortable means of travel. The approval of trailers for cruising speeds identical to buses without trailers (max. 100 km/h) entails no drawbacks in terms of travel times, increasing the attractiveness.

Body
Lightweight design based on a unitized body has gained acceptance as standard (integral construction). The body and base frames, which are firmly welded together, consist of pressed grid-type support elements and rectangular tubes (Figure 10).

In the case of the chassis structure, the body is positioned on a supporting ladder-type frame (similar to trucks). Except for mini- and midibuses, this design is not commonly encountered in Europe.

Chassis systems
The vertically or horizontally mounted engine drives the rear axle. Here the engine together with the transmission is not – in contrast to trucks – installed at the front

Figure 10: Unitized bus body
1 Body framework,
2 Pressed parts,
3 Base frame,
4 Rectangular tube,
5 Grid-type support elements.



and then connected with a long propeller shaft to the powered axle, but instead the engine/transmission unit is installed at the rear end of the bus. This also applies to articulated buses, in which the last axle is powered. Pneumatic suspension on all axles permits ride-level stabilization and a high degree of ride comfort. Intercity buses and tour buses are mainly equipped with independent suspension on their front axles. Disk brakes, frequently supported by retarders, are used on all axles.

Alternative drives
In view of the increasing sensitivity of users and operators of commercial vehicle, increasing endeavors are being made to produce alternative drives for commercial vehicles. The advantage of vehicles with alternative drives is that they comply with CO_2 legislation, avoid break-in restrictions, and satisfy the demands of society in general. These demands on the one hand relate to eliminating the dependence on petroleum fuels, whose days as fuels are numbered. They relate on the other hand to reducing the emission of CO_2 as a combustion product that contributes substantially to the greenhouse effect.

Both light utility vans and buses can be equipped for urban operation with a natural-gas, a battery-electric, or a fuel-cell drive. Likewise, battery drives are also available for heavy trucks in local goods and delivery traffic. A future development task is to come up with the best complete-vehicle concept for an electric truck and in so doing bring into line the general design of the battery with regard firstly to range, installation space, weight, and costs, and secondly to the demands of daily operation (charging cycle, charging time, useful load, ease of use).

Active and passive safety in commercial vehicles

Active safety comprises all the systems and measures which help to avoid an accident. Passive safety is intended to limit the consequences of accidents and to protect other road users. Systematic recording of accidents, accident tests with complete vehicles, and intensive computer optimization help to devise safety measures.

Active safety
A minimum of active safety systems is redefined on a regular basis by legislative bodies. Where the antilock braking system (ABS) was developed to the stage of series production for trucks back in 1980, it gained acceptance in all trucks at a later stage as the result of appropriate legal requirements. This also applies to driving-dynamics control (Electronic Stability Control, ESC) and the active brake assistant.

The availability of a comprehensive package of active-safety systems is a necessary precondition for the application of high automated or autonomous driving, a epoch-making change in truck operation.

Passive safety
Requirements
Generally, the effectiveness and strength of occupant restraint systems has to be demonstrated. Therefore, the dimensioning of commercial-vehicle body structures must take account of aspects, such as the strength and rigidity of seat-belt anchor points on the seats and of the related body structures (seat rails, floor, frame, etc.).

In the event of a collision, the driver's cab and the passenger cabin must maintain the amount of room necessary for occupant survival, while at the same time deceleration must not be excessive. Depending on vehicle design, there are a variety of solutions to this problem. In light utility vans, front-section design is energy-absorbing as in passenger cars. In spite of shorter deformation paths and higher levels of released energy, the physiologically permissible limits are not exceeded in virtually all passenger-car

crash-test standards (legal requirements and rating tests). In light utility vans, there must also be features which prevent injury to occupants by uncontrolled movement of the payload. The static and dynamic strength of such features (partitions, cages and nets, securing eyes) has to be demonstrated mathematically or by testing.

In the case of trucks, the side members extend up to the front fender and can absorb high linear forces. Such passive-safety measures are based on accident analyses and are intended to improve the structural design of the cab. Static and dynamic stress and impact tests at the front and rear of the cab, as well as on its roof, simulate the stresses involved in a frontal impact and in accidents in which the vehicle overturns, rolls over, or in which the cargo moves. They are described in the Regulation ECE R29 [3], but the passing of these tests as a condition of type approval is only required in a few European countries.

Statistical analyses have proved that the bus is the safest means of passenger transportation. Static roof-load tests and dynamic overturning tests provide evidence of body strength. The use of flame-retardant and self-extinguishing materials for the interior equipment of the vehicle minimizes the risk of fire.

As road traffic involves many different kinds of vehicles, collisions between light and heavy vehicles cannot be avoided 100 % without the use of active-safety systems or the communication of vehicles with each other (V2V, vehicle to vehicle; V2X, vehicle to infrastructure). As a result of the differences in vehicle weight, and incompatibility in terms of vehicle geometry and structural stiffness, the risk of injury in the lighter vehicle is greater.

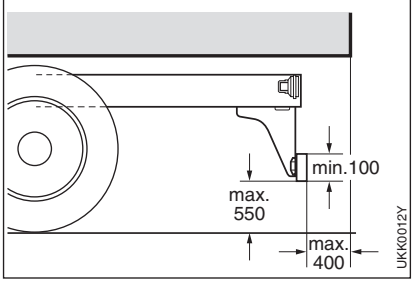
The change in speed Δv during a central plastic impact for front-end or rear-end collisions between two vehicles (vehicle 1 and vehicle 2) is defined as

$$\mu = \frac{m_2}{m_1}: \ \Delta v_1 = \frac{\mu v_r}{1+\mu}, \ \Delta v_2 = \frac{v_r}{1+\mu},$$

with the masses m_1 and m_2 of the vehicles involved and the relative speed v_r prior to the impact.

Side, front and rear underride guards help to reduce the danger of the lighter vehicle driving under the heavier vehicle. In other words, they serve to protect other road users (Figure 11).

Figure 11: Rear underride guard for trucks



References

[1] 96/53/EU: Council Directive 96/53/EC of 25 July 1996 laying down for certain road vehicles circulating within the Community the maximum authorised dimensions in national and international traffic and the maximum authorised weights in international traffic.

[2] 2015/719/EU: Directive (EU) 2015/719 of the European Parliament and of the Council of 29 April 2015 amending Council Directive 96/53/EC laying down for certain road vehicles circulating within the Community the maximum authorised dimensions in national and international traffic and the maximum authorised weights in international traffic (Text with EEA relevance).

[2] ECE R 29: Regulation No. 29; Uniform provisions concerning the approval of vehicles with regard to the protection of the occupants of the cab of a commercial vehicle: Revision 1.

[4] E. Hoepke, S. Breuer and others: Nutzfahrzeugtechnik. Verlag Springer Vieweg, 8th Ed., 2016.

Lighting equipment

Functions

Ever since the invention of the automobile, it has been accompanied by vehicle lighting. Initially candles, then petroleum and carbide lamps were used to provide lighting; today, these lamps would be classed as position or marker lamps. Only with the introduction of automotive electrics in the form of the Bosch alternator in the Adler vehicle (1913) were manufacturers able to introduce systems which generated good ranges and lived up to being called "headlamps". Further significant milestones have been

– the introduction of the asymmetrical low-beam pattern, characterized (RHD traffic) by an extended visual range along the right side of the road (1957),
– the introduction of new headlamp systems featuring complex geometrical configurations (PES, Poly-Ellipsoid System, free-form surfaces, facetted reflectors) offering efficiency-level improvements of up to 50 % (1985),
– the "Litronic" headlamp system with gas-discharge lamps (xenon lamps with luminous arc) which supplied more than twice the light generated by comparable halogen units (1990),
– adaptive front-lighting systems (AFS) with moving, dynamic PES modules (Poly-Ellipsoid System) or static, activated reflectors for turning (2003).

Lighting at the vehicle front end
The primary function of the headlamps at the vehicle front end is to illuminate the roadway so that the driver can detect traffic conditions and recognize any obstacles and hazards in good time. They also serve to identify and mark out the vehicle to oncoming traffic. The turn-signal lamps serve to show the driver's intention to change direction or to indicate a hazardous situation.

The headlamps and lights at the front end include the following:

– low-beam headlamps
– high-beam headlamps
– fog lamps
– auxiliary driving lamps
– turn-signal (direction-indicator) lamps
– parking lamps
– position/clearance lamps (for wide vehicles)
– daytime running lamps (if required by law in individual countries).

Lighting at the vehicle rear end
Lights are turned on at the vehicle's rear end in accordance with the lighting and weather conditions and indicate the vehicle's position. They also indicate how the vehicle is moving and in which direction, e.g., whether the brakes are applied or the driver is intending to change direction, or whether a hazardous situation exists. The reversing lamps illuminate the roadway while the vehicle is reversing.

The lamps and lights at the rear end include the following:

– stop lamps,
– tail lamps,
– rear fog warning lamps,
– turn-signal (direction-indicator) lamps,
– parking lamps,
– clearance lamps (for wide vehicles),
– reversing (backup) lamps,
– license-plate lamps.

Lighting in the vehicle interior
In the vehicle interior, priority over all other functions is given to the ease and reliability with which the switch elements can be reached and operated, and to provide the driver with sufficient information on the vehicle's operating states (while distracting him/her from driving as little as possible). These priorities dictate effectively illuminated instrument panels and discrete lighting for various control clusters (such as the sound and navigation systems), where they satisfy a prime requirement for relaxed and safe vehicle operation. Optical and acoustic signals must be prioritized according to their urgency and then relayed to the driver.

Regulations and equipment

Approval codes and symbols, Europe/ECE
Automotive lighting equipment is governed by national and international design and operating regulations, according to which the equipment in question must be manufactured and tested. For every type of lighting equipment, there is a special approval code and symbol which has to be legibly displayed on the equipment concerned. The preferred locations for approval codes and symbols are places that are directly visible when the hood is open, such as the lenses of headlamps and other lights, and the headlamp unit components. This also applies to approved replacement headlamps and lights.

If an item of equipment carries such an approval code/symbol, it has been tested by a technical inspectorate (e.g., in Germany the Lighting Technology Institute of Karlsruhe University) and approved by a licensing authority (in Germany the Federal Road Transport Office). All volume-production units which carry the approval code/symbol must conform in all respects to the type-approved unit. Examples of approval symbols:

(E1) ECE approval mark, [e1] EU approval mark.

The number 1 following each letter indicates the type-approval test carried out and the award of approval according to ECE (Economic Commission for Europe) Regulations for Germany with Europe-wide recognition. In Europe, installation of all automotive lighting and visual signaling equipment is governed not only by national guidelines but also by the higher European directives (ECE: whole of Europe, EU, New Zealand, Australia, South Africa and Japan). In the course of the ongoing union of Europe, the implementation regulations are being increasingly simplified by the harmonization of directives and legislation.

Right-hand-drive or left-hand-drive traffic
The ECE Regulations apply by analogy to driving on the right or left. The technical requirements for lighting are mirrored around the central perpendicular of the test screen (see Figure 4). According to the Vienna Global Treaty of 1968, when traveling in countries which drive on the other side of the road, all road users are obliged to adopt measures to prevent increased glare of oncoming traffic at night due to the asymmetrical light pattern. This can be achieved either by self-adhesive strips obtainable from the vehicle manufacturer or two-way switches in the headlamps (in the case of PES).

Regulations for the USA
In the USA, lighting equipment is governed by regulations that are very different from those in Europe. The principle of self-certification compels each manufacturer as an importer of lighting equipment to ensure, and in an emergency to furnish, proof that its products conform 100 % with the regulations of FMVSS 108 [27] (Federal Motor Vehicle Safety Standard) laid down in the Federal Register. There is therefore no type approval in the USA. The regulations of FMVSS 108 are partly based on the SAE industry standard (Society of Automotive Engineers).

Upgrading and conversion
Vehicles imported to Europe from other regions must be modified to comply with European directives. This applies in particular to the lighting equipment. Identical components available for the European market can be used as direct replacements. Other solutions, such as retail products or, in certain cases, retention of the original equipment, require an engineer's report. In Germany, Article 22a of the StVZO (Road Traffic Licensing Regulations) [1] requires "Approximation Certificates" for lighting equipment. Such certificates are issued by the Lighting Technology Institute of Karlsruhe University.

Subsequent alterations to type-approved headlamps, bases, and sockets invalidate the type approval and consequently the general operating license of the vehicle.

Light sources

With regard to automotive light sources, a distinction is made between thermoluminescent radiators (thermal radiators) and electroluminescent radiators. The outer electron shells of atoms of certain materials can absorb varying levels of energy by excitation (energy input). The transition from higher to lower levels may lead to the emission of electromagnetic radiation.

Thermoluminescent radiators
In the case of this type of light source, the energy level of the crystal system is increased by adding heat energy. Emission is continuous across a broad wavelength range. The total radiated power is proportional to the power of 4 of the absolute temperature (Stefan-Boltzmann law). The distribution-curve maximum is displaced to shorter wavelengths as temperature increases (Wien's displacement law) [38].

Filament lamps
Filament lamps, with tungsten filament (fusion temperature 3,660 K), are also thermal radiators. The evaporation of the tungsten and the resulting blackening of the bulb restrict the service life of this type of lamp.

Halogen lamps
A halogen filling (iodine or bromine) in the lamp allows the filament temperature to rise to close to the melting point of the tungsten. Close to the hot bulb wall, the evaporated tungsten combines with the filler gas to form tungsten halide. This is gaseous, light-transmitting and stable within a temperature range of 500 K to 1,700 K. It reaches the filament by means of convection, decomposes as a result of the high filament temperature, and forms an even tungsten deposit on the filament. In order to maintain this cycle, an external bulb temperature of approx. 300 °C is necessary. To achieve this, the bulb, made of fused silica (quartz), must surround the filament closely. A further advantage of this measure is that a higher filling pressure can be used, thereby providing additional resistance to tungsten evaporation.

Electroluminescent radiators
Gas-discharge lamps
Gas-discharge lamps are electroluminescent radiators and distinguished by their higher luminous efficiency. A gas discharge is maintained in an enclosed, gas-filled bulb by applying a voltage between two electrodes. The atoms of the emitted gas are excited by collisions between electrons and gas atoms. The atoms excited in the process give off their energy in the form of luminous radiation.

Examples of gas-discharge lamps are sodium-vapor lamps (street lighting), fluorescent lamps (interior illumination) and D lamps for motor-vehicle applications (Litronic).

Light-emitting diodes
A light-emitting diode or LED is an electroluminescent lamp. Thanks to their robustness, high energy efficiency, fast reaction time, and compact construction, LEDs are used as illuminators or displays in a wide range of applications. They are used in motor vehicles in the passenger compartment to provide illumination, as displays or for display backlighting. In the exterior areas LEDs are fitted in particular in auxiliary stop lamps and tail lamps. With a further increase in their luminous efficiency, they can increasingly be used in lamps and main functions at the front end of the vehicle.

Motor-vehicle bulbs

Replaceable filament bulbs for motor-vehicle lighting must be type-approved in accordance with ECE R37 [10], replaceable gas-discharge light sources in accordance with ECE R99 [19]. Other light sources which do not comply with these regulations (LEDs, neon tubes, special bulbs) are permitted, but can only be installed as a fixed component part of a lamp or as a "light-source module".

Bulbs complying with ECE R37 are generally available in 12 V versions, some bulbs also in 6 V and 24 V versions (Table 1). To help avoid mix-ups, different bulb types are identified by different base shapes. Bulbs of differing operating voltages are labeled with this voltage in order to avoid mix-ups in the case of identical bases. The bulb type suitable in each case must be indicated on the equipment.

A voltage increase in halogen lamps of 10 % results in a 70 % reduction in service life and a 30 % increase in luminous flux (Figure 1, [39]).

The luminous efficiency (lumens per watt) represents the bulb's photometrical efficiency relative to its power input. The luminous efficiency of vacuum bulbs is 10 to 18 lm/W. The higher luminous efficiency of halogen lamps (22 to 26 lm/W) is primarily a consequence of increasing the filament temperature. Gas-discharge bulbs provide a luminous efficiency in the order of 85 lm/W for substantial improvements in low-beam performance.

LEDs today achieve a luminous efficiency in the range of 50 lm/W (LEDs with high power consumption) or 100 lm/W (LEDs with low power consumption). An increase in luminous efficiency of up to 25 % is expected in the next few years.

Main-light functions

Low beam

The main light for driving at night is provided by low-beam headlamps. The creation of the characteristic light-dark boundary was one of the technological milestones in lighting technology.

The "dark above/bright below" distribution pattern resulting from the light-dark boundary furnishes acceptable visual ranges under all driving conditions. This configuration reduces glare, to which approaching traffic is exposed, within reasonable limits, and at the same time it supplies relatively high luminous intensity in the area below the light-dark boundary.

The light distribution pattern must combine maximum visual ranges with minimum glare effect. These demands are supplemented by other requirements affecting the area directly in front of the vehicle. For instance, the headlamps must provide assistance when cornering, i.e., the light distribution pattern must extend beyond the left and right-side extremities of the road surface.

With the asymmetrical low-beam pattern the light-dark boundary in the right area slopes upwards so that the right side of the road is better illuminated. This prevents oncoming traffic from being dazzled. This is mirror-inverted in left-hand-drive vehicles.

Figure 1: Influence of operating voltage on some data of halogen lamps

(Source: [39]).
L Service life (the dispersion width during operation with undervoltage is necessitated by the presence of halogen),
U Operating voltage, I Lamp current,
P Lamp power output, Φ Luminous flux.

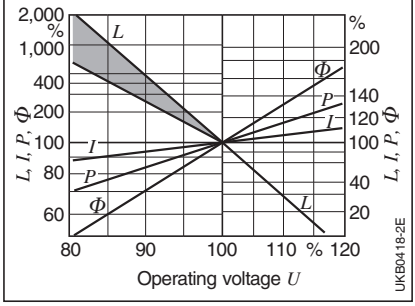


Table 1: Specifications for the main motor-vehicle bulbs (not including motorcycle bulbs)

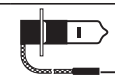
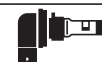
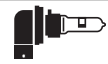
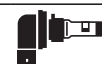
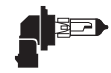
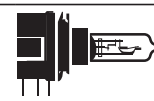
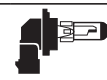
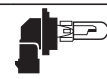
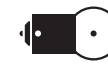
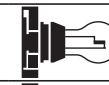
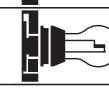
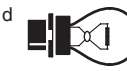
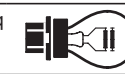
Application	Category	Voltage Nominal values V	Power Nominal values W	Luminous flux Reference values Lumen	Base type IEC	Illustration
Fog lamps, high beam, low beam in 4 HL	H1	6 12 24	55 55 70	1350 [2]) 1550 1900	P 14.5 e	
Fog lights, high beam	H3	6 12 24	55 55 70	1050 [2]) 1450 1750	PK 22s	
High beam, low beam	H4	12 24	60/55 75/70	1650/ 1000 [1]), [2]) 1900/1200	P 43 t - 38	
High beam, low beam in 4 HL, Fog lights	H7	12 24	55 70	1500 [2]) 1750	PX 26 d	
Fog lights, static cornering headlamps	H8	12	35	800	PGJ 19-1	
High beam	H9	12	65	2100	PGJ 19-5	
Low beam, fog lights	H11	12 24	55 70	1350 1600	PGJ 19-2	
Fog lights	H10	12	42	850	PY 20 d	
High beam, daytime running lamps	H15	12 24	55/15 60/20	260/1350 300/1500	PGJ 23t-1	
Low beam in 4 HL	HB4	12	51	1095	P 22 d	
High beam in 4 HL	HB3	12	60	1860	P 20 d	
Low beam, high beam	D1S	85 12 [5])	35 approx. 40 [5])	3200	PK 32 d-2	
Low beam, high beam	D2S	85 12 [5])	35 approx. 40 [5])	3200	P 32 d-2	
Low beam, high beam	D2R	85 12 [5])	35 approx. 40 [5])	2800	P 32 d-3	

Table 1 (contd.): Specifications for the main motor-vehicle bulbs (not including motorcycle bulbs)

Application	Category	Voltage Nominal values V	Power Nominal values W	Luminous flux Reference values Lumen	Base type IEC	Illustration
Stop lamps, turn-signal lamps, rear fog light, backup lamps	P 21 W PY 21 W[6])	6, 12, 24	21	460[3])	BA 15 s	
Stop lamp/ tail lamp	P 21/5 W	6 12 24	21/5[4]) 21/5 21/5	440/35[3]), [4]) 440/35[3]), [4]) 440/40[3])	BAY 15d	
Position lamp, tail lamp	R 5 W	6 12 24	5	50[3])	BA 15 s	
Tail lamp	R 10 W	6 12 24	10	125[3])	BA 15 s	
Daytime running lamps	P 13 W	12	13	250[3])	PG 18.5 d	
Stop lamp, turn signal	P 19 W PY 19 W	12 12	19 19	350[3]) 215[3])	PGU 20/1 PGU 20/2	
Rear fog light, backup lamp, front turn signal	P 24 W PY 24 W	12 12	24 24	500[3]) 300[3])	PGU 20/3 PGU 20/4	
Stop lamps, turn-signal lamps, rear fog light, backup lamps	P 27 W	12	27	475[3])	W 2.5 x 16 d	
Stop lamp/ tail lamp	P 27/7 W	12	27/7	475/36[3])	W 2.5 x 16 q	
License-plate lighting, tail lamp	C 5 W	6 12 24	5	45[3])	SV 8.5	
Position lamp	H 6 W	12	6	125	BAX 9 s	
Position lamp, license-plate lighting	W 5 W	6 12 24	5	50[3])	W 2.1 x 9.5 d	
Position lamp, license-plate lighting	W 3 W	6 12 24	3	22[3])	W 2.1 x 9.5 d	

[1]) High/low beam. [2]) Specifications at test voltage 6.3, 13.2 or 28.0 V.
[3]) Specifications at test voltage 6.75, 13.5 or. 28.0 V. [4]) Main/secondary filament.
[5]) With ballast unit. [6]) Yellow-light version.

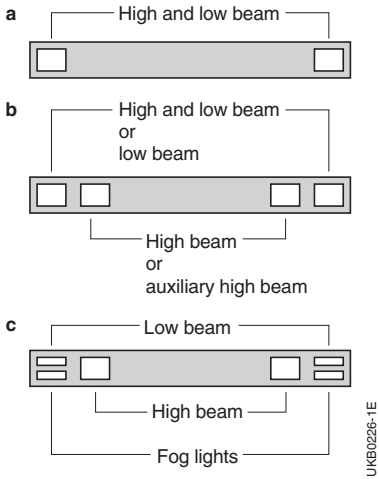
High beam
High-beam headlamps illuminate the road to the maximum range. This creates a high luminous intensity, depending on the distance, acting on all objects in the space reserved for traffic. High-beam headlamps are therefore only permitted when they do not dazzle oncoming traffic (glare).

The high traffic density on modern roads severely restricts the use of high-beam headlamps.

Installation and regulations
Designs
Global regulations mandate two headlamps for low beam and at least two (or the option of four) high-beam units for all dual-track vehicles. The light color is white.

Dual-headlamp system
The dual-headlamp system (Figure 2a) uses lamps with two light sources (halogen double-filament bulbs (H4), US sealed beam) for high and low beams using shared reflectors (see Reflection headlamps). In headlamps with gas-discharge bulbs the dual function is achieved by focusing or defocusing the xenon burner in a shared reflector (see Bi-Litronic "Reflection"). In bi-xenon projection systems a screen is moved out of or into the beam path (see Bi-Litronic "Projection").

Figure 2: Headlamp systems
a) Dual-headlamp system,
b) Quad-headlamp system,
c) Quad-headlamp system with additional fog lamps.



Quad-headlamp system
Two of the headlamps in a quad-headlamp system provide both high and low beam, while the second pair only provides high-beam illumination (Figure 2b). The light functions of projection and reflection systems can be used in any combination. The low-beam headlamps can additionally be combined with the fog lamps (Figure 2c).

Important definitions of device design
Grouped design
A single housing, but with different lenses and bulbs. Example:
– Multiple-compartment rear-lamp assemblies containing different individual light units.

Combined design
A single housing and bulb assembly with more than one lens. Example:
– Combined tail lamp and license-plate lamp.

Nested design
Common housing and lens, but with individual bulbs. Example:
– Headlamp assembly with nested position lamp.

Main-light functions for Europe

Regulations and directives for Europe
The most important regulations and directives are laid down in ECE R112 [20], ECE R113 [21], ECE R48 [12], 76/756/EEC [24], ECE R98 [18], and ECE R123 [23].
- ECE R112: Headlamps for asymmetrical low beam and/or high beam that are fitted with filament bulbs or LED modules (cars, buses, trucks).
- ECE R113: Headlamps for symmetrical low beam and/or high beam that are fitted with filament bulbs, gas-discharge light sources or LED modules (motorbicycles, motorcycles).
- ECE R48 and 76/756/EEC: for attachment and application.
- ECE R98: Headlamps with gas-discharge lamps as per ECE R99.
- ECE R123: Adaptive front-lighting systems (AFS) for motor vehicles.

The installation regulations described in the following refer to passenger cars.

Low beam
Regulations prescribe 2 white-light low-beam headlamps for multiple-track vehicles (Figure 3).

Automotive headlamp performance is subject to technical assessment and verification before they are put into volume production. Among the requirements are minimum luminous intensity, to ensure adequate road-surface visibility, and maximum intensity levels, to prevent glare (see measurement points and luminous intensities for headlamps, Figure 4 and Table 2).

Homologation testing is carried out under laboratory conditions using test lamps manufactured to more precise tolerances than those installed in production vehicles. The lamps are operated at the specified test luminous flux for each lamp category. The laboratory conditions apply across the board to all headlamps, but only take limited account of the specifics of individual vehicles, such as headlamp fitted height, vehicle power supply, and adjustment.

High beam
A minimum of two and a maximum of four headlamps are prescribed for the high-beam mode. The prescribed instrument-cluster high-beam indicator lamp is blue or yellow in color.

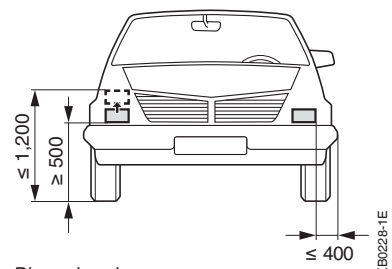
The high beam is usually generated by a light source located at the focal point of the reflector (Figure 5). This causes light to be reflected outward in the direction of the reflector axis. The maximum luminous intensity achievable by the high beam is largely a function of the reflector's illuminated area.

In quad-headlamp systems, in particular, roughly paraboloid high-beam reflectors can be replaced by units with complex geometrical configurations designed to supply a "superimposed" high-beam pattern. The calculations employed to design these units seek to achieve a high-beam distribution that harmonizes with the low-beam pattern (simultaneous activation). The pure high beam is "superimposed" as it were on the low-beam projection. The annoying overlap area close to the front of the vehicle is done away with in this case.

High-beam light distribution pattern is defined in the regulations and guidelines together with stipulations governing the low beams.

The maximum approved luminous intensity, a composite of the intensity ratings of all high-beam headlamps installed on the vehicle, is 430,000 cd. This value is indicated by identification codes located adjacent to the homologation code on

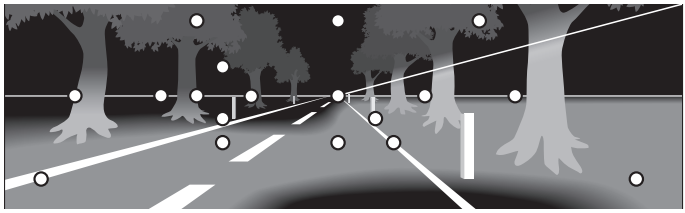
Figure 3: European headlamp system (low beam)



Dimensions in mm

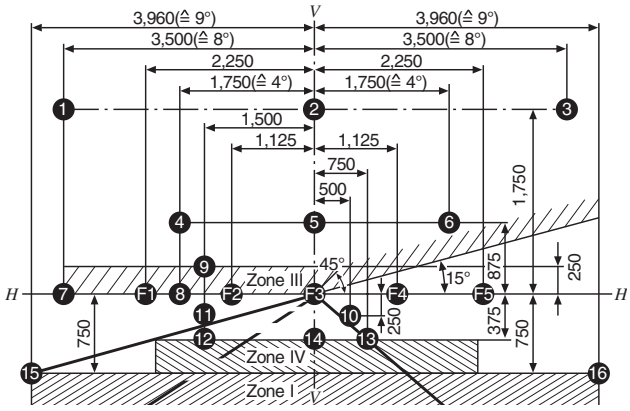
Figure 4: Luminous intensities of headlamps for Europe/ECE
a) Road perspective from driver's viewpoint,
b) Measurement point relative to road perspective as per ECE-R112.

a



UKB04419a-1E

b



UKB04419b-1E

Table 2: Measurement points and luminous intensities for headlamps

Low beam				High beam		
Measurement points on graphic			Luminous intensity	Measurement points		Luminous intensity
Figure no.	RHD traffic	LHD traffic	Class B [lx]	Figure no.	Point	Class B [lx]
1	8L/4U		≤ 0.7		E_{max}	48 < E < 240
2	V/4U		≤ 0.7			
3	8R/4U		≤ 0.7	F1	$E_{H\text{-}5.15°}$	> 6
4	4L/2U		≤ 0.7	F2	$E_{H\text{-}2.55°}$	> 24
5	V/2U		≤ 0.7	F3	E_{HV}[9]	≥ 0.8
6	4R/2U		≤ 0.7			E_{max}
7	8L/H	8R/H	≥ 0.1; ≤ 0.7	F4	$E_{H+2.55°}$	> 24
8	4L/H	4R/H	≥ 0.2; ≤ 0.7	F5	$E_{H+5.15°}$	> 6
9	B50L	B50R	≤ 0.4			
10	75R	75L	≥ 12			
11	75L	75R	≤ 12	For low beam: Total 1 + 2 + 3 ≥ 0.3 lx Total 4 + 5 + 6 ≥ 0.6 lx		
12	50L	50R	≤ 15			
13	50R	50L	≥ 12			
14	50V	50V	≥ 6			
15	25L	25R	≥ 2			
16	25R	25L	≥ 2			
Any point in zone III			≤ 0.7	[1] E is the current measured value in point 50R or 50L.		
Any point in zone IV			≥ 3			
Any point in zone I			≤ 2E[1]			

each headlamp. 430,000 cd corresponds to the figure 100. The luminous intensity of the high beam is indicated by the number, e.g., 25, stamped next to the round ECE approval mark. If these are the only headlamps on the vehicle (no auxiliary driving lamps), then the composite luminous intensity is in the range of 50/100 of 430,000 cd, i.e., 215,000 cd.

Auxiliary driving lamps
Auxiliary driving lamps are used to complement the effectiveness of the high beam in standard high-beam headlamps.

Auxiliary driving lamps are mounted and aimed in the same way as standard headlamps, and the underlying lighting technology is the same. Auxiliary driving lamps are also subject to the regulations governing maximum luminous intensity in vehicle lighting systems, according to which the sum of the reference numbers of all headlamps fitted to a vehicle must not exceed 100. For older headlamps without approval number, the number 10 is used for general assessment purposes.

Low- and high-beam switching
All high-beam headlamps must extinguish simultaneously when the low beams are switched on. Dimming (gradual deactivation) is permitted, with a maximum dimming period of 5 seconds. A 2-second response delay is required to prevent the dimming feature from activating when the headlamp flashers are used.

When the high-beam headlamps are switched on, the low-beam units may continue to operate (simultaneous operation). H4 bulbs are generally suitable for short periods of use with both filaments in operation.

Main-light functions for the USA

Regulations and directives
The national standard is the Federal Motor Vehicle Safety Standard (FMVSS) No.108 [27] and the SAE Ground Vehicle Lighting Standards Manual (Standards and Recommended Practices) to which it refers.

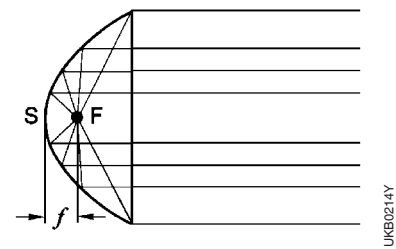
The regulations governing installation and control circuits for main headlamps are comparable in parts to those in Europe. Since 5/1/1997, however, headlamps with light-dark boundaries have also been authorized in the USA, but these require manual adjustment. It is now possible to develop headlamps which conform to legal requirements in both Europe and the USA.

As in Europe, dual- and quad-headlamp systems are used in the USA. The fitting and use of fog lamps and additional high-beam headlamps, however, are subject to a variety of, in some cases, widely diverging, local laws passed by the 50 individual states.

Up to 1983 only sealed-beam headlamps were permitted in the USA.

Low beam
The light-pattern requirements in America differ to a greater or lesser degree from the European system depending on the design type. In particular, the minimum glare levels are higher in the USA, and the maximum low-beam width is closer

Figure 5: Parabolic reflector
F Focal point, S Vertex of the parabola, f Focal length.



to the vehicle. The basic setting is generally higher (see measurement points in Figure 6 and Table 3).

High beam
The designs for high-beam headlamps are the same as in Europe. Differences exist in the required dispersion width of the light pattern, and there is a lower maximum figure on the axis of the high-beam headlamp.

Designs
Sealed-beam design
In this design, which is no longer used, the aluminized glass reflector and the lens must be sealed gas-tight on account of the light sources that are not encapsulated.

Figure 6: Luminous intensities of headlamps for the USA
a) Road perspective from driver's viewpoint,
b) Measurement point relative to road perspective as per FMVSS 108 (excerpt).

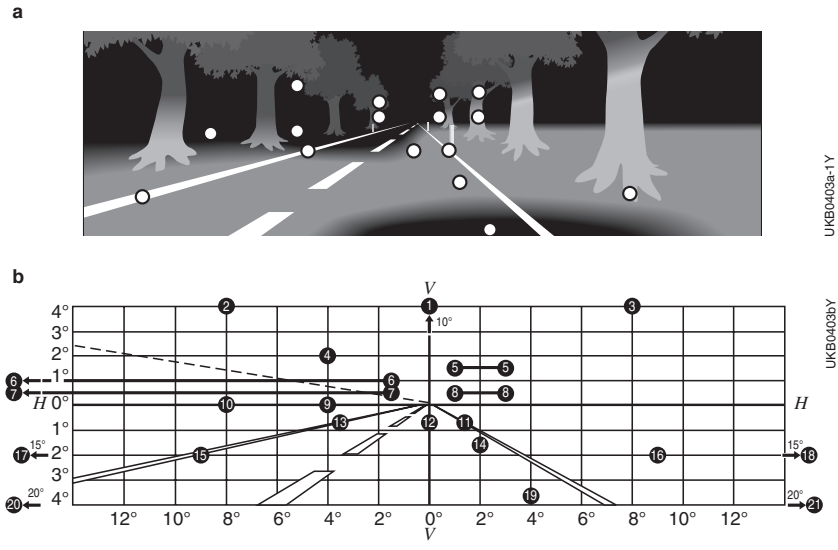


Table 3: Measurement points and luminous intensities for headlamps, low beam

Figure no.	Measurement points	Luminous intensity (cd)
1	10U-90U	≤ 125
2	4U, 8L	≥ 64
3	4U, 8R	≥ 64
4	2U, 4L	≥ 135
5	1.5U, 1R-3R	≥ 200
5	1.5U, 1R-R	≤ 1,400
6	1U, 1.5L-L	≤ 700
7	0.5U, 1.5L-L	≤ 1,000
8	0.5U, 1R-3R	≥ 500; ≤ 2,700
9	H, 4L	≥ 135
10	H, 8L	≥ 64

Figure no.	Measurement points	Luminous intensity (cd)
11	0.6D, 1.3R	≥ 10,000
12	0.86D, V	≥ 4,500
13	0.86D, 3.5L	≥ 1,800; ≤ 12,000
14	1.5D, 2R	≥ 15,000
15	2D, 9L	≥ 1,250
16	2D, 9R	≥ 1,250
17	2D, 15L	≥ 1,000
18	2D, 15R	≥ 1,000
19	4D, 4R	≥ 12,500
20	4D, 20L	≥ 300
21	4D, 20R	≥ 300

The whole unit is sealed and filled with an inert gas. If a filament burns through, the entire light source must be replaced. Units with halogen lamps are also available.

The limited range of available sealed-beam headlamps on the market severely restricted the freedom of design for vehicle front ends.

Vehicle headlamp aiming device (VHAD)
This design involves replaceable-bulb headlamps which are adjusted vertically with the aid of a spirit levels integrated in each headlamp, and horizontally by means of a system comprising a needle and dial. In fact, this is equivalent to "on-board aiming".

Headlamps for visual aim (VOL / VOR)
These systems have been in use since 1997. These are replaceable-bulb headlamps whose low beam has a light-dark boundary line which allows visual aiming of the headlamps (as is standard in Europe).

Either the left horizontal light-dark boundary (VOL, visual optical aim left; VOL mark on the headlamp) or, as is more often the case in the USA, the right horizontal light-dark boundary (VOR, visual optical aim right; VOR mark on the headlamp) is used. What is notable about the US systems is the position of the light-dark boundary, which is much closer to the horizon (inclination depending on type 0.4 to 0 %). This increases the potential dazzling risk of such systems.

There is no horizontal aiming with this type of headlamp.

Definitions and terms

Photometrical terms and definitions
Headlamp range
This is the distance at which the light beam continues to supply a specified luminous intensity – mostly the 1 lux line at the right side of the road (LHD traffic).

Geometric range of a headlamp
This is the distance to the horizontal portion of the light-dark boundary on the road surface (see Headlamp leveling control, Table 4). A low-beam inclination of 1 %, or 10 cm/10 m, results in a geometric range equal to 100 times the headlamp's fitted height (as measured between the center of the reflector and the road surface).

Visual range
The visual range is the distance at which an object (vehicle, object, etc.) within the luminance distribution of the human visual field is still visible.

The visual range is influenced by the shape, size, and reflectance of objects, the road-surface type, headlamp design and cleanliness, and the physiological condition of the driver's eyes. Due to the large number of influencing factors, it is not possible to quantify this range using precise numerical definitions. Under extremely unfavorable conditions (with RHD traffic, on the left side of a wet road surface) the visual range can fall to below 20 m. Under optimum conditions, it can extend outward to more than 100 m (with RHD traffic, on the right side of the road).

Disability: glare
This is the quantifiable reduction in visual performance that occurs in response to light sources emitting glare. An example would be the reduction in visual range that occurs as two vehicles approach one another.

Discomfort: glare
This condition occurs when a glare source induces discomfort without, however, causing an actual reduction in visual performance. Discomfort glare is assessed according to a scale defining different levels of comfort and discomfort.

Headlamp technology
Reflector focal length
Conventional reflectors for headlamps and other automotive lamps are usually parabolic in shape (Figure 5). The focal length f (distance between the vertex of the parabola and the focal point) is 15 to 40 mm.

Reflector illuminated area
This is the parallel projection of the entire reflector opening on a transverse plane. The standard reference plane is perpendicular to the vehicle's direction of travel.

Effective luminous flux, efficiency of a headlamp
The first of the above is the portion of the light source's luminous flux that is capable of supplying effective illumination via its reflective or refractive components (for instance, as projected on the road surface via the headlamp reflector). A reflector with a short focal length makes efficient use of the filament bulb and has high efficiency because the reflector extends outward to encompass the bulb, allowing it to convert a large proportion of the luminous flux into a useful light beam.

Angles of geometric visibility
These are angles that are defined relative to the axis of the lighting device, at which the illuminated area must be visible.

Technical design variations of headlamps

Components
Parabolic reflector
Reflectors direct the light from the light source either directly onto the road (reflection system) or into an intermediate plane which is further projected by a lens (projection system, see PES headlamps). The reflectors are made of plastic, die-cast metal or sheet steel.

Plastic reflectors are manufactured by injection molding (thermosetting plastics), which offers a considerably better precision of geometry reproduction than the deep-drawing of sheet-steel reflectors. The geometric tolerances that can be achieved lie in the range of 0.01 mm. The base material requires no corrosion-proofing treatment.

The die-cast metal used is usually aluminum, or occasionally magnesium. The advantages are high thermal resistance and the ability to produce shapes with a high degree of complexity (shaped bulb holders, screw holes and bosses).

The surfaces of thermosetting plastic and die-cast metal reflectors are given a smooth finish by spray painting or powder-based paint before a layer of aluminum 50 to 150 nm thick is applied. An even thinner transparent protective coat prevents the aluminum from oxidizing.

Sheet-steel reflectors are manufactured using deep-drawing and punching dies. A powder-based paint is then applied. This process hermetically seals the sheet steel and gives it excellent surface smoothness. The base coat created in this way is, like other reflectors, coated with aluminum.

Free-form reflectors
Furthermore, reflectors in stages and any desired facet distributions can also be implemented. The geometrical configurations of free-form reflectors are generated using complex mathematical calculations (HNS, Homogeneous Numerically Calculated Surface). Here, the low mean focal length f is defined relative to the distance between the reflector vertex and the center of the filament. Typical values range from 15 to 25 mm.

In the case of reflectors partitioned with steps or facets, each partition can be created with its own mean focal length *f*.

Diffusing lenses
A large proportion of contoured diffusing lenses is manufactured using high-purity glass (free of bubbles and streaks). During the lens molding process, high priority is given to surface quality in order to prevent undesirable upward light deflection in the final product – this would tend to create glare for oncoming traffic. The type and configuration of the lens prisms depend on the reflector and the desired light distribution pattern.

Cover lenses
The clear cover lenses used on modern headlamps are usually made of plastic. Besides reducing weight, plastic lenses provide other advantages for automotive applications, including greater freedom in headlamp and vehicle design. Since around 2007 multicolor plastic lenses (2-component lenses) have also been used in which the edge area is sprayed in a different color, usually black or gray. The advantage here is that the spray tools are designed in such a way that no inner slides are needed and therefore no separating lines are created on the visible surface. Light dispersion from the edge areas is also avoided.

There are several reasons why plastic cover lenses should not be cleaned with a dry cloth:
– Despite the scratch-proof coating, rubbing with a dry cloth can damage the surface of the lens.
– Rubbing with a dry cloth can produce an electrostatic charge in the lens, which can then allow dust to build up on the inside of the lens.

Conventional reflection headlamps
For conventional headlamp systems with virtually parabolic reflectors (Figures 5 and 7), the quality of the low beam increases in direct proportion to the size of the reflector. At the same time the geometric range increases as a function of installation height.

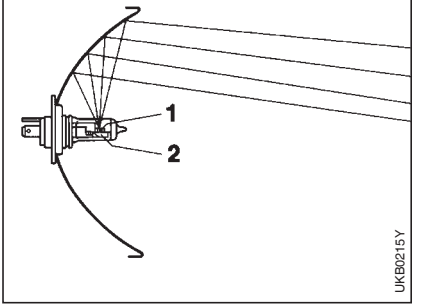
These factors must be balanced against the aerodynamic constraints according to which the vehicle's front-end profile must be kept as low as possible. Under these circumstances, increasing the size of the reflector results in wider headlamps.

Reflectors of a given size, but with different focal lengths, also perform differently. Shorter focal lengths are more efficient and produce wider light beams with better close-range and side illumination. This is of particular advantage during cornering.

Specially developed lighting programs (CAL, computer-aided lighting) enable the implementation of infinitely variable reflector shapes with non-parabolic sections and of faceted reflectors.

Headlamps with facet-type reflectors
In the case of facets, the reflector surface is partitioned and each individual segment is individually optimized. The important feature of surfaces with facet-type reflectors is that discontinuity and steps are permitted at all boundary surfaces of the partition. This results in freely shaped reflector surfaces with maximum homogeneity and side illumination (Figures 8 and 9).

Figure 7: Low beam (beam-path vertical section, H4 lamp)
1 Low-beam filament, 2 Bulb cover.

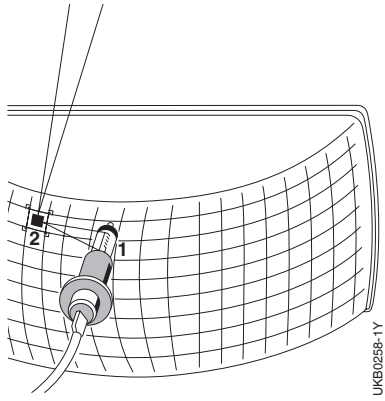


PES headlamps
The PES headlamp system (poly-ellipsoid system) employs imaging optics (Figure 10) and offers greater design scope than conventional headlamps. A lens aperture area with a diameter of only 40 to 70 mm allows the generation of light patterns previously only achievable with large-area headlamps. This result is obtained using an elliptical (CAL designed) reflector in combination with optical projection technology. A screen reflected with the objective projects precisely defined light-dark boundaries. Depending on specific individual requirements, these transitions can be defined as sudden or gradual intensity shifts, making it possible to obtain any geometry required.

PES headlamps can be combined with conventional high beams, position lamps and PES fog lamps to form lighting-strip units in which the entire headlamp is no higher than approximately 80 mm.

In PES headlamps, the beam path can be configured in such a way that the surroundings of the objective are also used in the signal image. This enlargement of the signal image is used above all with small objective diameters to reduce the psychological glare for the oncoming traffic. The additional surface can also be made into a lens, a partially vapor-deposited screen or a design feature with illuminated round or rectangular gaps, or illuminated three-dimensional objects.

Figure 8: Free-form or facet-type reflector
Filament pattern reflection with mirror optics.
1 Filament, 2 Mirror optics.



Xenon headlamps
Functioning principle
The headlamp system with a xenon gas-discharge lamp as its central component (Figure 11) generates high illumination-intensity levels with minimal frontal-area requirements, making it ideal for vehicles with aerodynamic styling with exceptional c_w values. In contrast with the conventional filament bulb, light is generated by a plasma discharge inside a burner the size of a cherry stone (Figure 12).

Figure 9: Facet-type reflector
a) Vertically partitioned,
b) Radially and vertically partitioned.

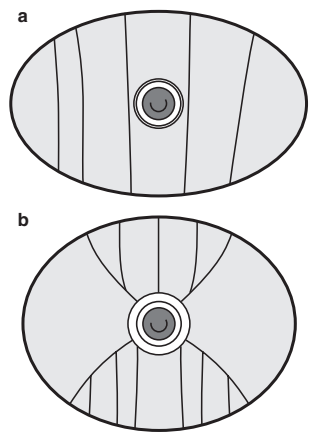
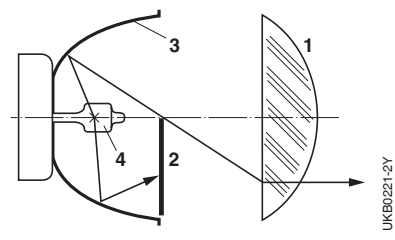


Figure 10: PES reflector (optical principle)
1 Objective, 2 Screen, 3 Reflector,
4 Lamp.



Performance
The arc of the 35W xenon bulb D2S delivers, when compared with the H7 bulb, twice the luminous flux with a higher color temperature (4200 K), i.e., the color of the light is very similar to natural sunlight. A D5S bulb with a power rating of only 25 W still delivers three times the luminous flux of an H7 bulb with the same high color temperature as a 35W xenon bulb. Maximum luminous efficiency, corresponding to approx. 90 lm/W, is available as soon as the quartz element reaches its operating temperature of more than 900 °C. Brief high-power operation at currents of up to 2.6 A (continuous operation: approx. 0.4 A) can be used to obtain "instant light". 2000 hours of service life are sufficient for the average required total duration of operation in passenger cars. As no sudden failure occurs as in the case of a filament, diagnosis and replacement in good time are possible.

Categories
At present gas-discharge lamps with the type designations D1 and D2 and since 2012 only the type designations D3 and D4 are used in the power-rating range of 35 W. In the D3 and D4 families dosing of the heavy metal mercury (approximately 1 mg) can be dispensed with. These lamps are distinguished by lower lamp voltage, a different plasma composition, and different arc geometries. The same applies to the new categories of 25W gas-discharge lamps D5, D6 and D8. The electronic control units for the individual lamp types are generally developed for a specific design type and are not universally interchangeable.

The D2 and D4-series automotive gas-discharge lamps feature high-voltage-proof bases and UV glass shielding elements. On the D1 and D3-series models, the high-voltage electronics necessary for operation are also integrated in the lamp

Figure 12: Gas-discharge lamp D5S
1 UV glass shielding bulb,
2 Lead-in insulator,
3 Discharge chamber (burner),
4 Electrodes,
5 Bulb base,
6 Integrated ignition and control unit.

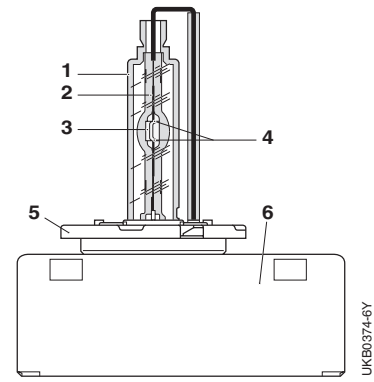
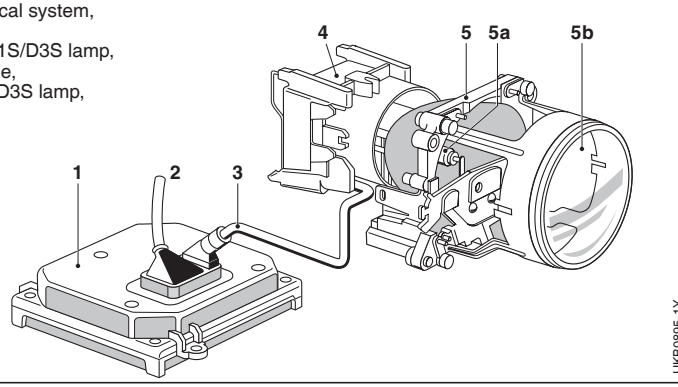


Figure 11: System components for PES-design Litronic headlamp
1 ECU,
2 To vehicle electrical system,
3 Shielded cable,
4 Ignition unit of D1S/D3S lamp,
5 Projection module,
5a Burner of D1S/D3S lamp,
5b Lens.



base. Category D5 lamps are a special case – these lamps have the advantage that both the ignition unit and the control unit are already integrated in the base of the gas-discharge lamp.

All systems feature two subcategories: S-lamps for projection-system headlamps and R-lamps for reflection headlamps with an integrated light shield for producing the light/dark boundary comparable with the bulb cover used for halogen H4 low beams. At present the D1S and D3S lamps are the type most widely used.

Electronic ballast unit

An integral part of the Litronic headlamp (contraction of Light and Electronics) is the electronic ballast unit responsible for activating and monitoring the lamp (Figure 13). Its functions include ignition of the gas discharge (voltage 10 to 20 kV), controlled power supply during the warm-up phase when the lamp is cold, and demand-oriented supply during stationary operation.

The system furnishes largely consistent levels of illumination (i.e., elimination of luminous-flux changes) by compensating for fluctuations in the vehicle system voltage. If the lamp goes out (for instance, due to a momentary voltage drop in the vehicle electrical system), re-ignition is spontaneous and automatic.

The electronic ballast unit responds to defects (such as a damaged lamp) by interrupting the power supply to help avoid injury in the event of contact.

Figure 14: Light distribution pattern on the road (comparison)

a) Halogen H4 lamp,
b) Litronic PES D2S lamp.

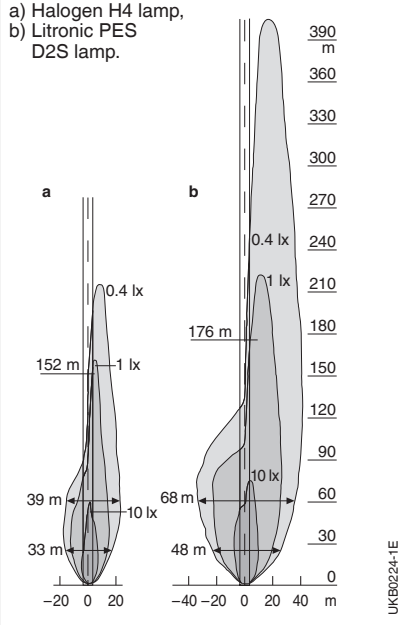
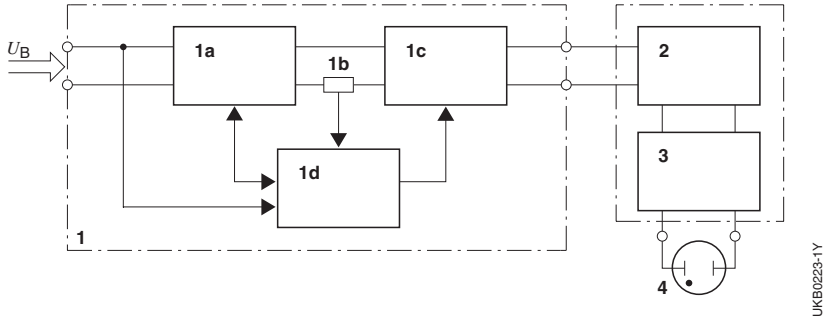


Figure 13: Electronic ballast unit for 400-Hz AC supply and pulse triggering of the bulb

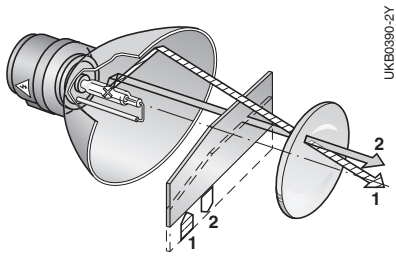
1 ECU, 1a DC/DC converter, 1b Shunt, 1c DC/AC converter, 1d Microprocessor, 2 Ignition unit, 3 Lamp socket, 4 D2S lamp, U_B Battery voltage.



Light distribution
The xenon light emitted by Litronic headlamps produces a broad carpet of light in front of the vehicle combined with a long range (Figure 14). This has made it possible to achieve substantially wider road-illumination patterns for lighting the edges of bends and wide roads as effectively as a halogen unit illuminates straight stretches of road. The driver enjoys substantial improvements in both visibility and orientation in difficult driving conditions and bad weather.

Headlamp leveling control
In accordance with ECE Regulation 48 [12], Litronic headlamps with 35W gas-discharge lamps are combined with automatic headlamp leveling control and headlamp cleaning systems. These requirements do not apply to headlamps with 25W lamps on account of the lower luminous flux of the lamps, i.e., such headlamps may also – like halogen headlamps – be operated with manual headlamp leveling control. This makes it easier to switch from halogen to 25W gas-discharge headlamps in the aftermarket sector.

Figure 15: Bi-Litronic "Projection"
1 Low beam, 2 High beam.



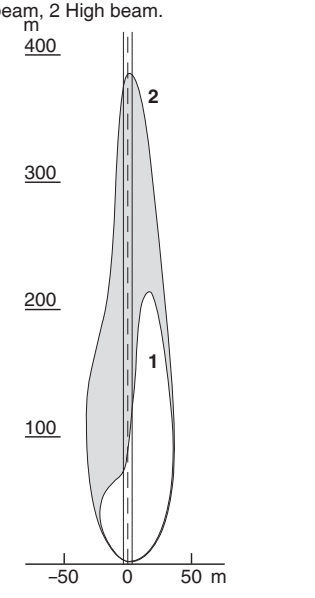
Bi-Litronic
Bi-Litronic systems permit both the low and high beams to be generated by the arc of a single gas-discharge lamp.

Bi-Litronic "Projection" (bi-xenon)
The Bi-Litronic "Projection" system is based on a PES Litronic headlamp (Figure 15). For the high beam the light shield for creating the light/dark boundary (for the low beam) is moved out of the light beam. With lens diameters of 70 mm, Bi-Litronic "Projection" allows the present highly compact headlamp design with combined high and low beam and at the same time outstanding luminous efficiency (Figure 16).

The primary advantage of Bi-Litronic "Projection" is the xenon light for high-beam operation.

Bi-Litronic "Reflection"
Both mono- and bi-xenon systems use a single DxR lamp for both headlamp functions.

Figure 16: Light distribution patterns of Bi-Litronic
1 Low beam, 2 High beam.



In the bi-xenon version, when the high-/low-beam switch is operated, an electromechanical actuator moves the gas-discharge lamp in the reflector into the appropriate position for producing the high or low beam as required (Figure 17).

LED headlamps

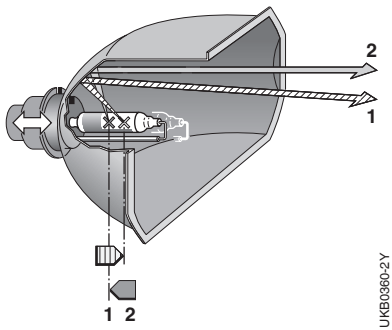
Potential for reducing energy consumption

LEDs are increasingly being used as economical alternatives in reducing CO_2 emissions and fuel consumption. The energy consumption of the main-light functions will play a significant role in future low-energy vehicles. Xenon and LED alternatives offer the levels of optimized consumption and improved road safety called for by the EU.

Today's LED systems already consume, depending on performance (luminous flux, range, side illumination), much less energy than halogen bulbs. Power consumption currently fluctuates between 28 W and 50 W per headlamp. When compared with bulb power output of approx. 65 W (for 13.2 V), this means potential savings of 30 to 70 W per vehicle.

One projection system and two reflection elements are used for the low-beam spot. The light of three multichip LEDs each with two LEDs is concentrated by three primary optical elements and imaged by a projection lens. The optical system incorporates a screen to guarantee the quality of the light-dark boundary. A reflector is positioned both above and below the lens (Figure 18).

Figure 17: Bi-Litronic "Reflection"
1 Low beam, 2 High beam.

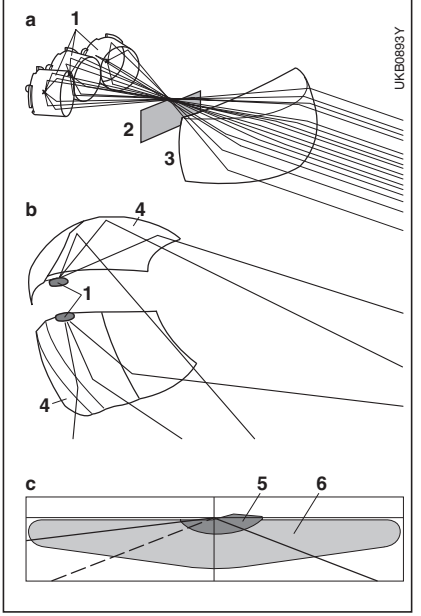


The optical efficiency of the LED low-beam headlamp is roughly 45 %. Compared with this, the efficiency of a bi-xenon system is around 33 %. This can be explained by the nature of LEDs, which give off their light only to the half-space and do not, like conventional light sources, illuminate the entire space. Because of the higher efficiency of an LED system, less luminous flux is needed in the LEDs to deliver the same luminous flux to the road with one LED low-beam headlamp.

The better LEDs will become in future, the less power will be output by the control unit, the light performance always being kept at a constantly high level.

The first LED headlamps on the market also show the increasing role of design

Figure 18: Basic design of an LED low-beam headlamp
Interaction of projection and reflection systems.
a) Projection system for the spot,
b) Reflection system for the basic light,
c) Total light distribution.
1 LEDs, 2 Light shield for light-dark boundary,
3 Projection system, 4 Reflector, 5 Spot,
6 Basic light.



elements in headlamps. In 2008, for example, a headlamp received an internationally renowned design award for the first time.

Around 75 % of total worldwide mileage is completed during daytime hours. Great importance is therefore also attached to the energy consumption of the daytime driving lights function. A typical LED daytime driving light has an energy consumption of 14 W (0.36 g CO_2/km) per vehicle. Use of vehicle lights during the day consumes up to 300 W (7.86 g CO_2/km) (low-beam headlamps, tail lamps, position lamps, number-plate lamps, switch and instrument lighting).

Laser light
Functions in the headlamp whose light source is based on a blue laser diode (see Laser LED) are called laser light. Similarly to white LEDs the blue laser radiation is converted with a phosphor conversion into a white wavelength spectrum and diffused. The advantage of the laser-light source including converter is the five to seven times higher luminance compared with the LEDs currently used. The higher luminance is used to generate light distributions with high luminous intensities with smaller reflectors. This can be used for example to produce an auxiliary high beam which roughly doubles the range on the road. The higher luminous intensities on the road lead to improved, earlier detection of objects at great distances and represent a significant increase in safety.

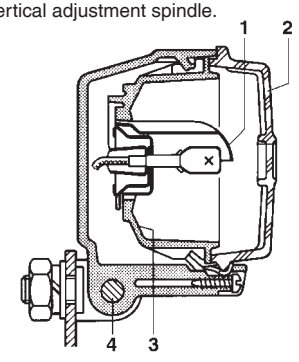
Additional headlamps

Fog lamps
Fog lamps (white light) are intended to improve orientation in fog, snow, heavy rain, and dust. A beam of light with a particularly high side dispersion is generated for this purpose. This ensures that the side of the road which is close to the vehicle is particularly well illuminated. The brightness levels achieved on close objects are significantly higher. Unlike the usually dark road surface far ahead of the vehicle, these high brightness levels help drivers to find their bearings better despite the poor weather conditions.

Designs
DIY and dealer-installed fog lamps are designed as individual projection units in their own housings. These are installed either upright on the fender, or suspended from it (Figure 19). Stylistic and aerodynamic considerations have led to increased use of integrated fog lamps, designed either for installation within body openings or included as a component within a larger light assembly (with adjustable reflectors when the fog lamps are combined with the main headlamps).

Present fog lamps produce white light. There is no substantive evidence that yellow lamps provide any physiological benefits. A fog lamp's effectiveness depends on the size of the illuminated area and the focal length of the reflector. Assuming the

Figure 19: Fog lamp (upright mounting)
1 Bulb cover, 2 Lens, 3 Reflector, 4 Vertical adjustment spindle.



same illuminated area and focal length, the differences between round and rectangular fog lamps from the technical viewpoint are negligible.

Regulations
Design is governed by ECE R19 [8], installation by ECE R48 [12] (and StVZO (Road Traffic Licensing Regulations) Article 52 [3] in Germany); two white or yellow fog lamps are permitted. The control circuit for switching the fog lamps must be independent of the high- and low-beam circuits. In Germany, the StVZO allows fog-lamp installation in positions more than 400 mm from the widest point of the vehicle's width, provided they are wired so that they can only be switched on when the low-beam headlamps are on (Figure 20).

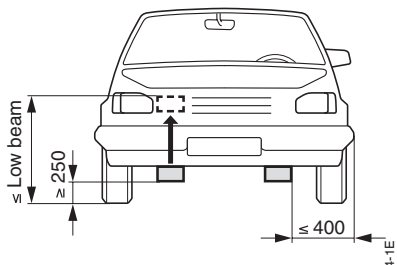
Paraboloid with diffusing lenses
A parabolic reflector featuring a light source located at the focal point reflects light along a parallel axis (as with the high-beam headlamp), and the lens extends this beam to form a horizontal band (Figure 19). A special bulb cover prevents the beam from being projected upward.

Free-form technology
Calculation methods, such as CAL (computer-aided lighting), can be used to design reflector shapes in such a way that they scatter light directly (i.e., without optical lens contouring) and also generate (without separate shading) a sharp light-dark boundary. The fact that the lamp features pronounced envelopment of the bulb leads to an extremely high volume of light combined with maximum dispersion width (Figure 21).

PES fog lamps
This technology minimizes reflected glare in fog. The screen, the image of which is projected onto the road surface by the lens, provides a light-dark boundary with minimum upward light dispersion.

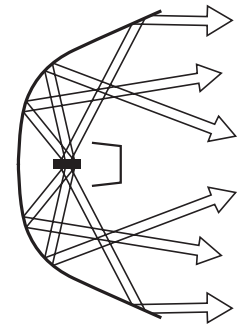
Innovations
The technical-functional importance of the fog lamp has been pushed into the background slightly by the introduction of powerful halogen, xenon and AFS systems. The emphasis is still on the demand for improved lighting in traffic situations with adverse ambient conditions. This can be solved by separate fog lamps of the type used up to now (as per ECE R19 [8]), by the poor-weather light function of AFS systems (as per ECE R123 [23]), or by combinations of the two functions.

Figure 20: Fog lamp (positioning)



Dimensions in mm

Figure 21: Fog lamp with free-form reflector (horizontal section)



Light functions

Cornering headlamps
Cornering headlamps have been approved for public use since the beginning of 2003. Whereas previously only high-beam headlamps were allowed to turn in response to changes in steering angle (1960s Citroën DS), swiveling low-beam headlamps are now also permitted (dynamic cornering headlamps or adaptive headlamps) or a supplementary light source (static cornering headlamps). This provides for a larger visual range on winding roads.

Static cornering headlamps
Static cornering headlamps are used mainly to illuminate areas close to the side of the vehicle (switchbacks, turning maneuvers). For this purpose, activating additional reflector elements is generally the most effective way.

Dynamic cornering headlamps
Dynamic cornering headlamps are used to illuminate the changing course of the road, e.g., on winding overland highways (Figure 22).

In contrast with the directly linked swiveling action of the cornering headlamps of the 1960s, modern "high-end systems" electronically control the rate of swivel and the swivel angle in response to the vehicle's speed. This optimizes "harmonization" between the headlamps and the vehicle attitude, and eliminates "jerky" headlamp movements. Headlamp positioning is performed by a positioner unit (stepping motor) which moves the basic or low-beam module or the reflector elements in response to changes in steering-wheel angle or the steering angle of the front wheels (Figure 23). Sensors detect those movements to prevent glare for oncoming traffic by means of "failsafe"

Figure 23: Cornering-headlamp modules
1 Supporting frame, 2 Mounting frame,
3 Drive motor for horizontal rotation,
4 Bi-Litronic-PES.

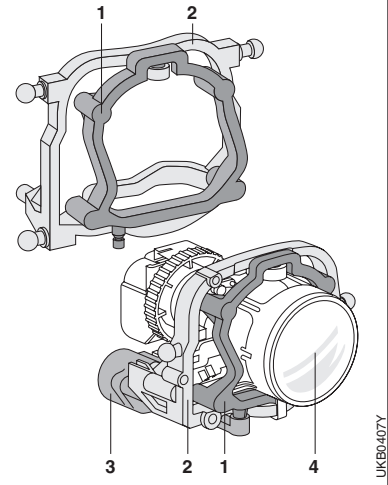
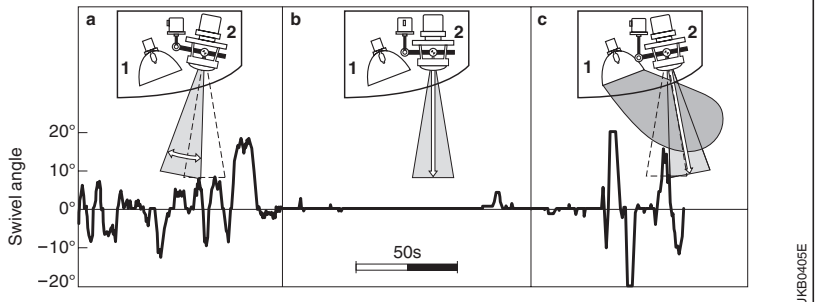


Figure 22: Switch and aiming strategy of turning and basic modules of a static/dynamic cornering headlamp (left side)
a) "Highway/cornering" position, b) "Expressway" position, c) "City/turning" position.
1 Turning module, 2 Basic module.



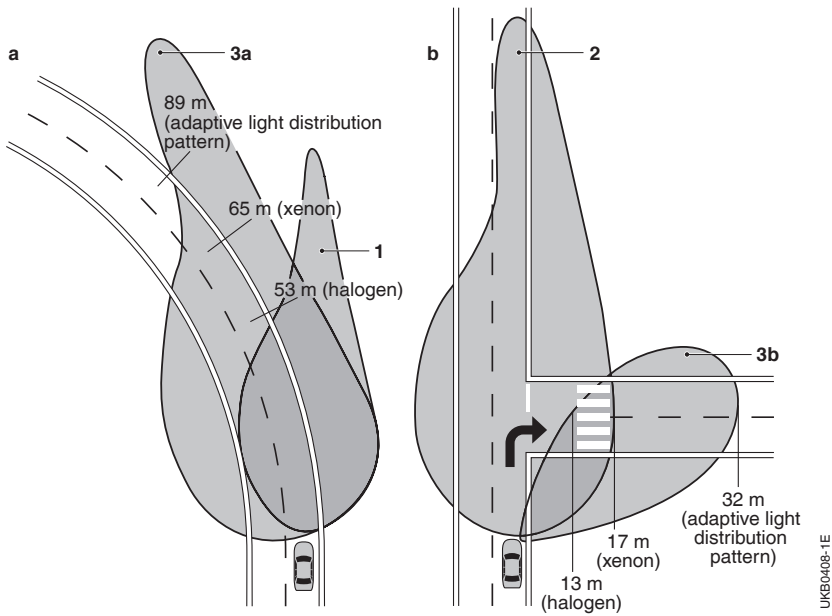
algorithms. General legal requirements specify that the headlamp beam may only be turned as far as the center line of the road at a distance of approx. 70 m in front of the vehicle to prevent glare for oncoming vehicles.

Road safety and driving convenience
The introduction of dynamic cornering headlamps is a significant improvement in the safety and convenience of driving at night (Figure 24). Compared with conventional low-beam headlamps, improvements in visual range of approximately 70 % are achieved, representing an extra 1.6 seconds of travel time. With cornering headlamps, a motorist can assess hazards better and start braking sooner. As a result, the severity of an accident can be significantly reduced. Static cornering headlamps double the visual range for turning maneuvers.

AFS functions
Expressway beam
For special driving situations, altered light patterns (AFS, adaptive front-lighting system) have been developed to enable a better view for the driver for each driving state. In the development of the expressway beam (Figure 25), special attention was paid to achieving a better range for the driver without dazzling oncoming traffic to an excessive degree. The increase in the detection distance to up to 150 m enables an extension of driving time up to the detected object of approx. 2 seconds (comparison at 100 km/h with halogen headlamps). This enables the driver to better assess a critical situation and, possibly, to initiate braking much earlier.

Poor-weather light
In the case of the poor-weather light beam, the special focus is on improving

Figure 24: Measurable improvement of visibility for driver with adaptive light distribution pattern of cornering headlamps
a) Left-hand bend, dynamic cornering headlamps, b) Right turn, static cornering headlamps.
1 Halogen headlamps, 2 Xenon headlamps,
3a Adaptive light distribution pattern with dynamic cornering headlamps,
3b Adaptive light distribution pattern with static cornering headlamps.



the optical guidance on the road. Especially the zones around the sides of the road are better illuminated.

Most variants of the poor-weather light feature movement of the left cornering-headlamp module by 8° to the side and simultaneously a slight lowering or activation of the static cornering lights. This provides for very wide illumination of the road and the edge of the road. In future, component elements, for example the elements responsible for widened side illumination, will be activated sequentially. Control parameters are, for example, steering-angle information and direction-indicator actuation. The individual segments are then "quasidynamically" activated.

Light functions and driver-assistance systems

The introduction of video technology in automotive engineering makes it possible to implement camera-based headlamp functions as well. When the position of an oncoming vehicle is identified by the camera, the headlamp or the AFS system can adapt the range of the driving lights in such a way that it is increased for large distances and reduced for smaller distances (dynamic range function). This ensures optimum illumination without dazzling oncoming traffic.

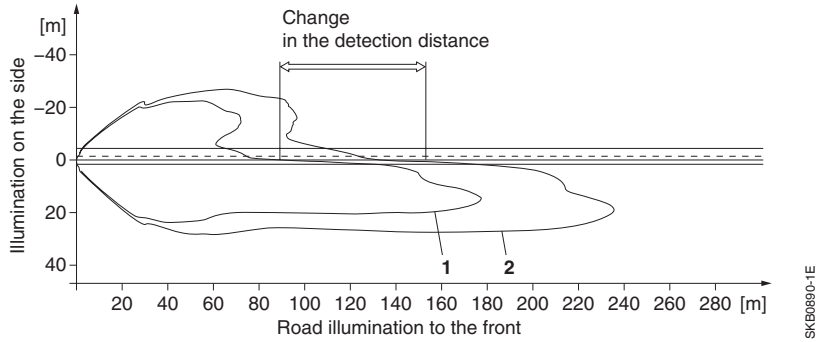
Glare-free high beam

"Glare-free high beam" is a light function which is classed as part of the high beam but in contrast to the conventional high beam can also be activated in traffic situations with preceding or oncoming traffic. This function is referred to in the international context and in the statutory provisions as Adaptive Driving Beam (ADB). Activation requires a camera and image analysis which determines the positions of the other road users quickly and precisely. Shadows in the high-beam distribution are specifically controlled with this information (Figure 26, see also High-beam assistant).

Dynamic system

This type is based on the dynamic cornering headlamps (see Figure 23). Here the left headlamp generates a partial high-beam distribution with a vertical light/dark boundary to the right side and the right headlamp generates a right partial high-beam distribution (Figure 26). In the "normal" high beam both the partial high-beam distributions are superimposed in the middle to achieve high luminous intensities in the central range. To prevent other road users from being dazzled, a shadow is generated in the corresponding position and width with the dynamic movement of both sides.

Figure 25: Metrological-functional connection between expressway beam and highway beam
1 Road illumination with highway beam, 2 Road illumination with expressway beam.

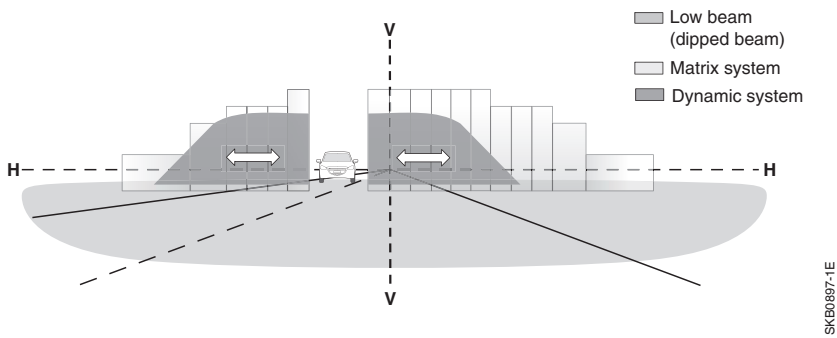


Matrix system
In realizing the glare-free high beam as a matrix system the light distribution is made up typically of 12 to 18 strips. Individual LEDs or LED arrays with individually switchable chips are used as light sources and the specific attributes such as compact shape, rapid switch-on and switch-off, and dimmability are used. Each strip is formed by an LED chip. The shadow adapted to the traffic situation is formed by the specific switching-off of individual LEDs or groups of LEDs. For "normal" high-beam distribution all the LEDs are operated and a balanced light distribution is created by means of a dimming profile.

Matrix headlamps
Three-row modules with 84 pixels are available in some vehicles as a further development of the one-row matrix systems for glare-free high beam. The shadows for fading out other road users can be positioned even more precisely with the finer pixel resolution (Figure 27). Further shadows can additionally be created for example to fade out road signs.

Matrix in the low beam
The essential difference from the one-row systems is that with this type part of the low beam is also formed with the matrix module. This enables even the dynamic cornering headlamps – without mechanical actuators – to be implemented by purely electronic means via activation of the LEDs. In addition, the specific dimmability of areas of the light distribution can be used in adverse weather light to reduce the dazzling of oncoming traffic via the reflection on the wet road.

Figure 26: Glare-free high beam (1st generation)
Schematic diagram.
V Vertical,
H Horizontal.



Installation and regulations for signal lamps

These types of lamp are intended to facilitate recognition of the vehicle, and to alert other road users to any intended or present changes in direction or motion. Uniform, distinctive colors in the red, yellow or white color range are prescribed for these lamps to denote their application. White, yellow and red lights are used to mark the positions of the vehicle at the front, the sides and the rear respectively. Stop lamps and rear fog warning lamps are also red. Yellow lights are used in most applications for turn-signal lamps. Only in the USA are red turn-signal lamps also permissible at the rear.

As projected along the reference axis, minimum and maximum luminous intensities for all lamps must remain within a range calculated to guarantee signal recognition without, however, causing glare nuisance for other road users.

Turn-signal lamps and hazard-warning flashers
(as per ECE R6 [6]).
ECE R48 and 76/756/EEC specify Group 1 (front), Group 2 (rear), and Group 5 (side) turn-signal lamps for vehicles with three or more wheels. For motorcycles, Group 2 turn-signal lamps are sufficient.

The lamps are electrically monitored. A function indicator is required inside the vehicle. The dashboard-mounted monitoring lamp may be in any color desired.

The flashing frequency is 90 ± 30 cycles per minute with a relative illumination period of 30 to 80 %. Light must be emitted within 1.5 s of lamp switch-on. All the turn-signal lamps on one side of the vehicle must flash synchronously. If one lamp fails, the remaining lamps must continue to generate visible light.

All the turn-signal lamps flash synchronously in hazard-warning mode, also operational when the vehicle is stopped. An operation indicator is mandatory.

Two lamps each (color: yellow) are prescribed for the front, the rear and the side turn-signal lamps. In the USA the color red or yellow is permitted for the rear and side turn-signal lamps (SAE J588, Nov. 1984, [33]).

Design regulations
The design requirements for Europe are set out in the ECE Regulations R6, R7, R23, R38 and R87 [6, 7, 9, 11, 16], and the installation requirements in ECE R48 [12] (Figures 28, 29 and 30).

In the USA, FMVSS 108 specifies the number, location, and color of signal lamps. The design and technical lighting requirements are defined in the relevant SAE standards.

Figure 27: Matrix headlamps (2nd generation)
Schematic diagram.

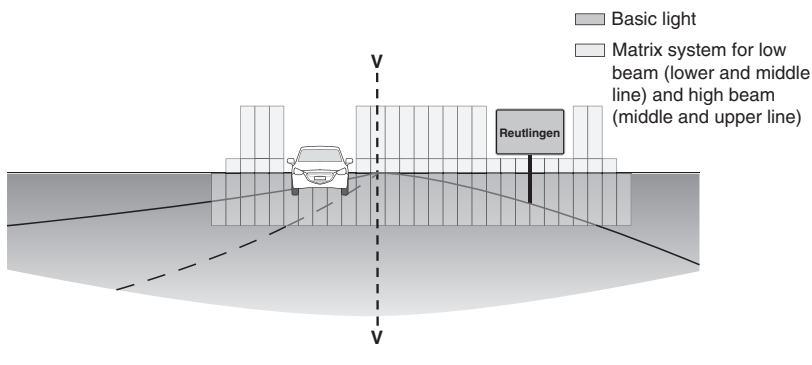


Figure 28: Positioning of front turn-signal lamps as per ECE
Dimensions in mm.
[1]) Less then 2100 mm if vehicle body type prohibits compliance with regulations on maximum height.

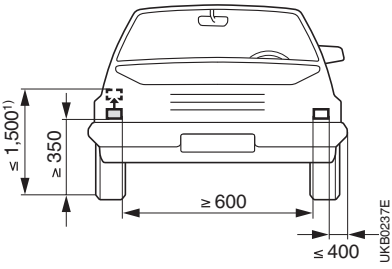


Figure 29: Positioning of rear turn-signal lamps as per ECE
Dimensions in mm. Height and width same as front turn-signal lamp.
1 Tail lamp.

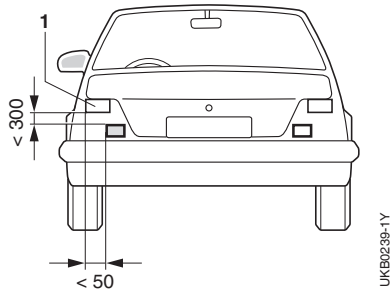
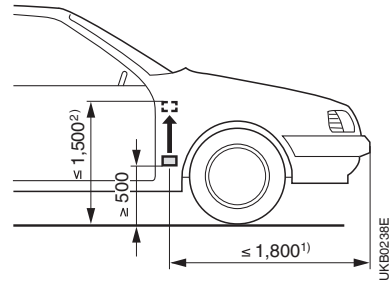


Figure 30: Positioning of side turn-signal lamps as per ECE
Dimensions in mm.
If the type of vehicle body does not permit adherence to the maximum dimensions:
[1]) or 2500 mm.
[2]) or 2300 mm.



Hazard-warning and turn-signal flashers for vehicles without trailer
The electronic hazard-warning and turn-signal flashers include a pulse generator designed to switch on the lamps via a relay and a current-controlled monitoring circuit to modify the flashing frequency in response to bulb failure. The turn-signal control stalk controls the turn signals, whereas the hazard flashers are operated using a separate switch.

Hazard-warning and turn-signal flashers for vehicles with trailer
This type of hazard-warning and turn-signal flasher differs from those employed on vehicles without trailers in the way that the function of the turn-signal lamps is controlled when they flash to indicate a change in direction.

Single-circuit monitoring
The tractor and trailer share a single monitoring circuit designed to activate the two indicator lamps at the flashing frequency. This type of unit cannot be used to localize lamp malfunctions. The flashing frequency remains constant.

Dual-circuit monitoring
Tractor and trailer are equipped with separate monitoring circuits. The malfunction is localized by means of the indicator lamp. The flashing frequency remains constant.

Tail and position lamps
(as per ECE R7 [7]).
According to ECE R48 and 76/756/EEC, vehicle and trailer combinations wider than 1,600 mm require position lamps (facing forwards). Tail lamps (at rear) are mandatory equipment on vehicles of all widths. Vehicles wider than 2,100 mm (e.g., trucks) must also be equipped with clearance lamps visible from the front and rear.

Position lamps
Two white-light position lamps are stipulated. The regulations set out in SAE J222, Dec. 1970, [29] apply in the USA.

Tail lamps
Two red tail lamps are stipulated. When the tail and stop lamps are combined in

a nested design, the luminous-intensity ratio for the individual functions must be at least 1:5. Tail lamps must operate together with the position lamps.

The regulations set out in SAE J585, Feb. 2008, [30] apply in the USA.

Clearance lamps
(as per ECE R7 [7]).
Vehicles wider than 2,100 mm require two white lamps facing forward, and two red lamps facing to the rear. They must be positioned as far outward and as high as possible.

The regulations set out in SAE J592e, [34] apply in the USA.

Side-marker lamps
(as per ECE R91 [17]).
According to ECE R48, vehicles of any length exceeding 6 m must have yellow side-marker lamps (SML) except on vehicles with cab and chassis only.

Type SM1 side-marker lamps may be used on vehicles of all categories; type SM2 side-marker lamps, on the other hand, may only be used on cars.

The regulations set out in SAE J592e] apply in the USA.

Rear reflectors
(as per ECE R3 [4]).
According to ECE R48, two red, non-triangular, rear reflectors are required on motor vehicles (one on motorcycles and motorbicycles).

Additional reflective items (red reflective tape) are permitted if they do not impair the function of the legally required lighting and signaling equipment.

Two colorless, non-triangular front reflectors are required on trailers and on vehicles on which all forward-facing lamps with reflectors are concealed (e.g., retractable headlamps). They are permitted on all other types of vehicle.

Yellow, non-triangular, side reflectors are required on all vehicles with a length exceeding 6 m, and on all trailers. These are permitted on vehicles shorter than 6 m.

Two red triangular rear reflectors are required on trailers, but are banned on motor vehicles.

There may be no light fitted inside the triangle.

The regulations set out in SAE J594, Feb. 2010, [36] apply in the USA.

Parking lamps
(as per ECE R77 [15]).
ECE R48 permits either two parking lamps at the front and rear, or one parking lamp on each side. The prescribed colors are white facing forwards and red facing rearwards. Yellow may also be used at the rear if the parking lamps have been designed as single units together with the side turn-signal lamps.

The parking lamps must be designed to operate even when no other vehicle lights (headlamps) are on. The parking-lamp function is usually assumed by the tail and position lamps.

The regulations set out in SAE J222, Dec. 1970, [29] apply in the USA.

License-plate lamps
(as per ECE R4 [5]).
According to ECE R48, the rear license plate must be illuminated so as to be legible at a distance of 25 m at night.

Across the complete license-plate, luminance must be at least 2.5 cd/m^2. The luminance gradient of $2 \times B_{min}$/cm should not be exceeded between any of the test points distributed across the surface of the license plate. B_{min} is defined as the smallest luminance measured at the test point.

The regulations set out in SAE J587, Oct. 1981, [32] apply in the USA.

As an alternative to the license-plate lamp, self-illuminating license plates will also be permitted in future.

Stop lamps
(as per ECE R7 [7]).
According to ECE R48, all cars must be fitted with two type S1 or S2 stop lamps, and one type S3 stop lamp, colored red in each case (Figure 31).

When a nested design with stop and tail lamps is used, the luminous-intensity ratio between individual functions must be at least 5 : 1.

The Category S3 stop lamp (central high-level stop lamp) must not be a combined unit incorporating any other lamp.

The regulations set out in SAE J586, Feb. 1984, [31] and SAE J186, Nov. 1982, [28] apply in the USA.

Rear fog warning lamps
(as per ECE R38 [11]).
For EU/ECE countries, ECE R48 prescribes one or two red rear fog warning lamps for all new vehicles. They must be distanced at least 100 mm from the stop lamp (Figure 32).

The visible illuminated area along the reference axis may not exceed 140 cm². The circuit must be designed to ensure that the rear fog warning lamp operates only in conjunction with the low beam, high beam and/or front fog lamp. It must also be possible to switch off the fog warning lamps independently of the front fog lamps.

Rear fog warning lamps may only be used if the visual range due to fog < 50 m, as rear fog warning lamps have a high luminous intensity that can severely dazzle drivers at the rear when there is a clear view. The required indicator lamp is yellow in color.

Reversing (backup) lamps
(as per ECE R23 [9]).
According to ECE R48, one or two white reversing lamps are permitted (Figure 33). The switching circuit must be designed to ensure that the reversing lamps operate only when reverse gear is engaged and the ignition on.

The regulations set out in SAE J593c, Feb. 1968, [35] apply in the USA.

Daytime running lamps and daytime driving lights
In the vehicle type approval the fitting and use of daytime running lamps has according to ECE R87 ([16], Figure 34) been mandatory in Europe since February 2011 for passenger cars and light commercial vehicles; since August 2012 this has also applied to all other vehicle categories.

Daytime running lamps must be automatically switched on on the vehicle when

Figure 31: Positioning of stop lamps as per ECE
Dimensions in mm.
1 Central high-level stop lamp (Category S3),
2 Two stop lamps (categories S1 and S2).
[1]) ≥ 400 mm if width < 1300 mm,
[2]) ≤ 2100 mm if compliance with max. height not possible, or
[3]) ≤ 150 mm below bottom edge of rear window,
[4]) However, the lower edge of the central high-level stop lamp must be higher than the upper edge of the main stop lamps.

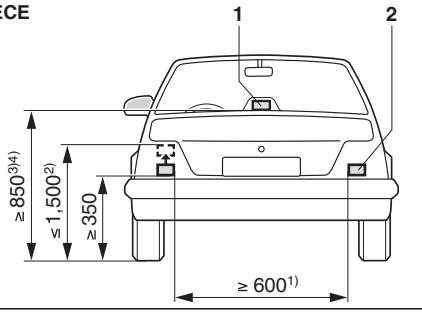
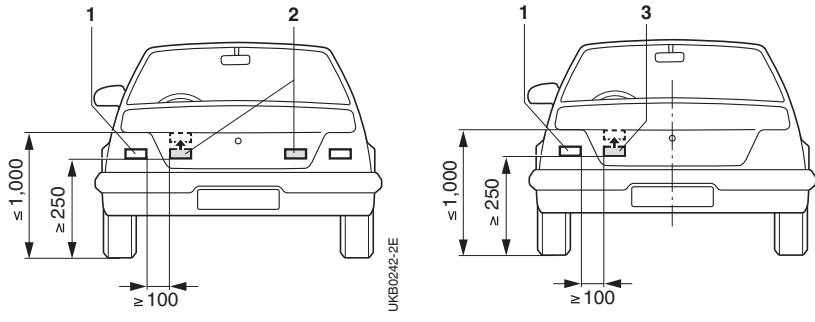


Figure 32: Positioning of rear fog warning lamps as per ECE
Dimensions in mm.
1 Stop lamp, 2 Two rear fog warning lamps, 3 One rear fog warning lamp for driving on right.



the engine is started and remain lit until either the front headlamps or fog lamps are switched on or the engine is turned off.

The regulations set out in FMVSS108 [27] apply in the USA.

Front cornering lamps
(as per ECE R119 [22]).
Two cornering lamps are permitted on the front of the vehicle which must radiate light at least 60° to the outside of the vehicle. They are used in cornering maneuvers up to a maximum speed of 40 km/h to illuminate better the destination road (e.g., side roads or garage entrances) which is usually inadequately lit by the regular headlamps. The front cornering lamps are activated via the turn-signal control stalk or when the steering wheel is turned.

Front cornering lamps typically have the same light distribution as static cornering headlamps. The switching conditions, however, are different. For the most part the same reflector is used for both the static cornering headlamps and the front cornering lamps.

The regulations set out in SAE J852, Apr. 2001, [37] apply in the USA.

Identification lamps
According to ECE R65 [14], identification lamps must be visible from any direction, and create the impression of flashing. The flashing frequency is 2 to 5 Hz. Blue identification lamps are intended for installation on official vehicles. Yellow identification lamps are designed to warn of dangers or the transport of dangerous freight.

Technical design variations for lamps

Light color
Depending on their application (e.g., stop lamps, turn-signal lamps, or rear fog warning lamps), motor-vehicle lamps must display uniform, distinctive colors in the red or yellow color range. These colors are defined in specific ranges of a standardized color scale (color location).

Since white light is composed of various colors, filters can be used to weaken or filter out the emission of undesirable spectral ranges (colors) completely. The color-filter functions may be performed by either the tinted lenses of the lamp, or the colored glass bulb of the lamp (e.g., yellow bulb in turn-signal lamps with clear lens).

Figure 33: Positioning of reversing lamps as per ECE
Dimensions in mm.
Quantity: 1 or 2.

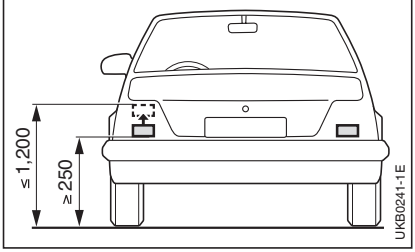
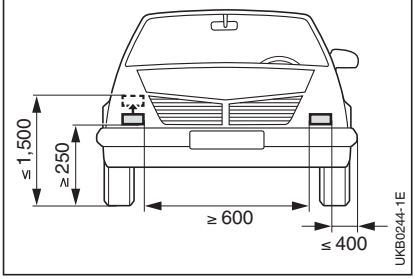


Figure 34: Positioning of daytime running lamps as per ECE
Dimensions in mm.



Filter technology can also be used to design lamp lenses in such a way that, when the lamp is switched off, the color is matched to the vehicle's paintwork. Nevertheless, existing homologation regulations are complied with when the lamp is on. Color locations have been laid down in the EU/ECE region. They correspond to a wavelength of approximately 592 nm for "yellow/orange" turn-signal lamps, for example, and to a wavelength of approximately 625 nm for "red" brake and tail lamps.

Luminous intensities

As projected along different axes, minimum and maximum luminous intensities for all lamps must remain within a range calculated to guarantee signal recognition without, however, causing glare nuisance for other road users. The same percentile basis ("unified spatial light distribution pattern") is used for most lamps. With regard to the value along the reference axis, luminous-intensity levels to the sides and above and below may be lower (Figure 35). However, these values may, depending on the installation height or for special lamps (e.g., daytime running lamps), differ from this basis.

Lamps with Fresnel optics

The light from the bulb is projected directly (without diversion) on the lens, which uses Fresnel optics to shape the beam in the desired manner (Figure 36). This concept is a highly cost-effective solution in that no vapor-deposited surface is needed. The drawbacks are low efficiency and the limited vehicle-design options.

Lamps with reflector optics

Lamps with approximately parabolic reflectors or stepped reflectors direct the light from the bulb in an axial beam, and shape the beam by means of the optical dispersion elements in the lens (Figure 37).

Lamps with free-form reflectors achieve the required beam spread or light pattern, completely or partially, by directing the light by means of a reflector. The outer lens can thus be designed as a clear lens, or supplemented by cylindrical lenses in the horizontal or vertical direction.

Designs combining both the above principles have also been successfully employed. The principle of the free-form lamp with Fresnel cap combines excellent luminous efficiency with a variety of possible stylistic implementations (Figure 38). It is essentially the reflector that shapes the light pattern. The Fresnel cap improves luminous efficiency by diverting a proportion of the light, which would not otherwise contribute to the function of the lamp, in the desired direction.

Mainly lamps with free-form reflectors are used in modern vehicle designs. These lamps offer the best opportunities of adaptation to the body shape and therefore to the available structural space,

Figure 35: ECE "lamp" test screen
Schematic diagram of spatial light distribution pattern (values in %).

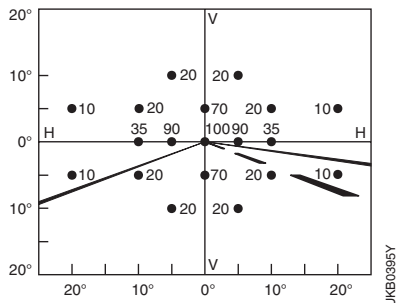
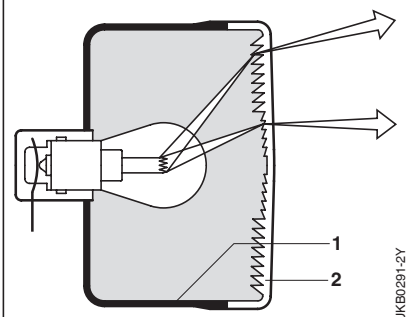


Figure 36: Lamp with Fresnel optics (principle)
1 Housing,
2 Lens with Fresnel optics.



and at the same time of meeting stylistic requirements.

Lamps with light-emitting diodes

Light-emitting diodes (LEDs) are increasingly used as the light source for lamps. LEDs have been used for years as standard in auxiliary stop lamps. The crucial factor here is the design freedom to deliver a slimline design with a number of light sources (Figures 39 and 40). In addition, LEDs in stop lamps provide an added safety aspect. An LED reaches maximum light output in less than 1 ms, whereas filament bulbs require about 200 ms to reach their nominal luminous flux. This means that LEDs emit the brake signal sooner, and this in turn reduces reaction time.

With their improved efficiency with regard to luminous flux, luminance, thermal performance and mechanical design, LEDs can also be used for functions with higher lighting requirements.

When it comes to being used as light sources in vehicle lights, LEDs differ crucially from filament and halogen lamps in that they cannot be operated directly from

Figure 37: Lamp with reflector optics (principle)
1 Reflector,
2 Lens with cylindrical dispersion elements.

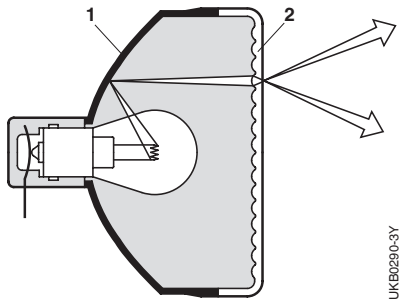


Figure 38: Free-form lamp with Fresnel cap (principle)
1 Reflector, 2 Fresnel cap,
3 Clear lens.

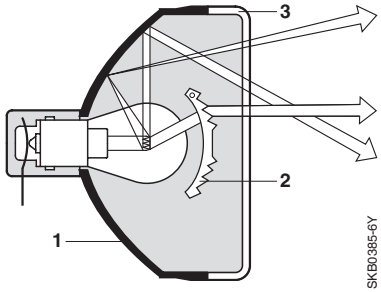


Figure 39: LED lamp with Fresnel optics (principle)
a) Vertical section, b) Overhead view.
1 Fresnel optics, 2 LED.

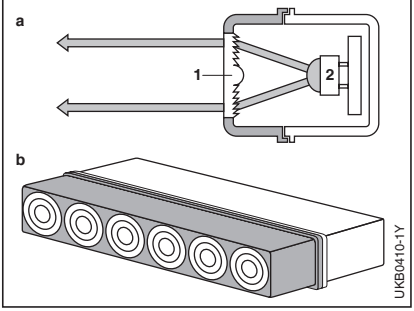
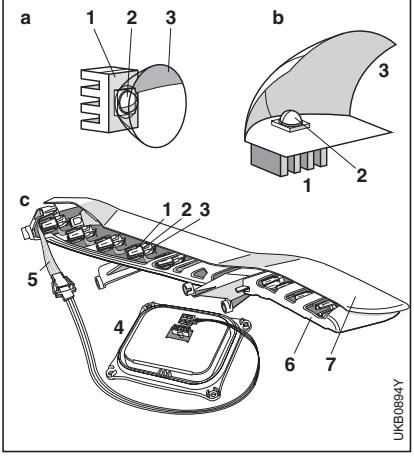


Figure 40: LED lamp
a) With attachment optics, b) With reflector,
c) Example of a daytime running lamp.
1 Heat sink, 2 LED, 3 Optics, 4 ECU,
5 Flexboard, 6 Holder, 7 Cover.



the vehicle electrical system. They require both a defined voltage, which ranges depending on the semiconductor material between 2.2 V and 3.6 V, and a defined current, by means of which the luminous intensity is adjusted. Resistor solutions can be used for activating functions with very low photometric requirements. Most applications require linear controllers or DC/DC converters. The electronics can be either integrated as individual electronics in the lamp or in a self-contained ECU, which is either mounted to the lamp or in the vehicle.

Lamps with fiber-optic technology

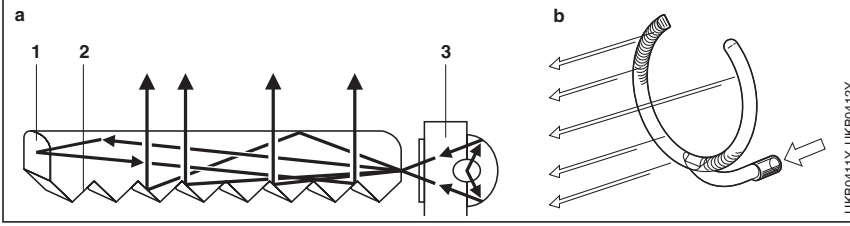
By using optical fibers/waveguides, the light source can be separated from the light emission point. In order to obtain the desired light pattern, special beam launchers in the optical waveguides, or optically active elements on or in front of the optical waveguide are required. Filament lamps can be used as light sources, but the high infrared levels are a drawback of these lamps. This necessitates the use of heat-resistant materials such as glass or heat shields. When using LEDs as "cold" light sources, it is possible to launch directly into transparent plastics such as PC (polycarbonate) or PMMA (polymethyl methacrylate). Lamps with fiber-optic technology are used primarily to produce stylistic elements such as narrow strips or thin rings (Figure 41).

Adaptive rear-lighting systems

Until now, rear-lighting functions were produced with a one-level circuit. Depending on the version and design, this resulted in a fixed luminous intensity that had to lie within the legal limit values to guarantee minimum visibility.

A large number of sensors (brightness, dirt, visual range, wetness, etc.) mean that the vehicle is nowadays able to ascertain the ambient parameters and lighting conditions more exactly. So that the optimal visibility (e.g., sufficient luminous intensity without excessive glare) can be achieved, rear light functions will be able to vary the luminous intensity they emit depending on the ambient conditions detected around the vehicle. For example, a stop lamp can be operated with high luminous intensity during strong sunlight and with lower values at night in order to ensure optimal detection and attribution of the driver's actions (Figure 42).

Figure 41: Lamp using fiber-optic technology (principle)
a) Design and function, b) Example.
1 Optical waveguide, 2 Beam launcher,
3 Lamp (with reflector and heat shield) or LED.



Convenience lamps

Lights for use when the vehicle is stationary are increasingly being fitted. Typical applications are the illumination of areas immediately outside the doors when the doors are open or closed, lights that identify the vehicle outline in the underbody area, or lights which identify the door handles.

Combinations of such lights with position lamps or fog lamps are referred to as "coming home" functions, which are activated when the doors are unlocked, for instance.

Headlamp leveling control

Headlamp adjustment for low and high beams

The correct adjustment of vehicle headlamps is an essential factor in nighttime road safety for both the driver of the vehicle concerned and for oncoming vehicles. If the beam is set only a fraction too low, there will be a substantial reduction in headlamp geometric range (Table 4). If the beam is set only a fraction too high, oncoming traffic will suffer much greater glare.

Installation and regulations

Headlamp leveling, Europe

Since 1/1/1998, an automatic or manual headlamp leveling control (beam-height adjustment) device has been mandatory in Europe for all first-time vehicle registrations, except in cases where other equipment (e.g., hydraulic suspension leveling) guarantees that light-beam inclination will remain within the prescribed tolerances.

Figure 42: Permitted light intensity for adaptive tail lamps

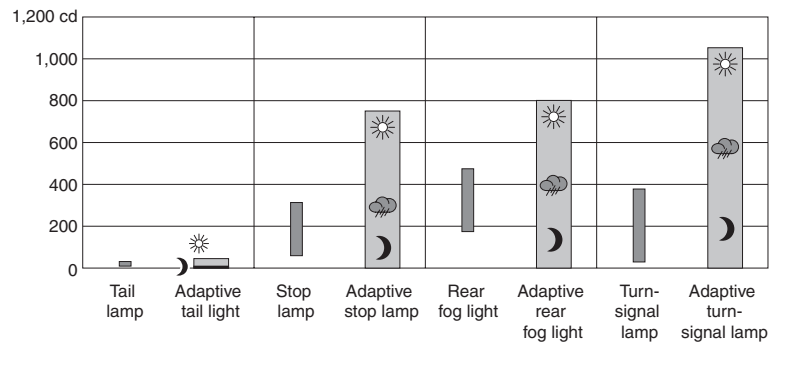


Table 4: Geometric range of the horizontal component of the low beam's light-dark boundary
Headlamp installed at height of 65 cm.

Inclination of the light-dark boundary (1 % = 10 cm / 10 m)	%	1.0	1.5	2.0	2.5	3.0
Adjustment dimension e	cm	10	15	20	25	30
Geometric range for the horizontal portion of the light-dark boundary	m	65.0	43.3	32.5	26.0	21.7

Although this equipment is not mandatory in other countries, its use is permitted.

Automatic headlamp leveling control must be designed to compensate for vehicle laden states by lowering or raising the low beam by between 5 cm/10 m (0.5%) and 25 cm/10 m (2.5%).

A manual headlamp leveling control device is operated from the driver's seat and must incorporate a detent position in the base setting; beam adjustment is also performed at this position. Units with infinitely variable and graduated control must both feature visible markings in the vicinity of the hand switch for vehicle load conditions that require vertical aiming.

Technical design variations
All design variations employ an adjustment mechanism to provide vertical adjustment of the headlamp reflector (housing design) (Figure 43). Manually operated units employ a switch near the driver's seat to control the setting (Figure 44), while automatic units rely on level sensors on the vehicle axles to monitor suspension spring compression, and relay the proportional signals to the adjustment mechanisms.

Hydromechanical systems
Hydromechanical systems operate by transmitting a fluid through the connecting hoses between the hand switch (or level sensor) and the adjustment mechanisms. The degree of adjustment corresponds to the quantity of pumped fluid.

Vacuum systems
With vacuum systems, the hand switch (or level sensor) regulates vacuum from the intake manifold, and transmits it to the adjustment mechanisms to achieve varying degrees of adjustment.

Electrical systems
Electric systems employ electric gear motors as the adjustment mechanisms. They are switched either by switches in the vehicle or by axle sensors.

Headlamp adjustment

Headlamp adjustment, Europe
Regulations and procedure
Correct adjustment of motor-vehicle headlamps should ensure the best possible illumination of the roadway at the same time as minimizing glare for other traffic. The EU Directives (and therefore the requirements of Article 50 of the StVZO [2] (Road Traffic Licensing Regulations) in Germany) specify the required horizontal and vertical alignment of the headlamp beams. Low-beam glare is regarded as eliminated if the luminous intensity is not more than 1 lux at a distance of 25 m in front of each headlamp when the beam

Figure 43: Automatic headlamp leveling control (principle)
1 Adjustment mechanism,
2 Processing unit,
3 Level sensors.

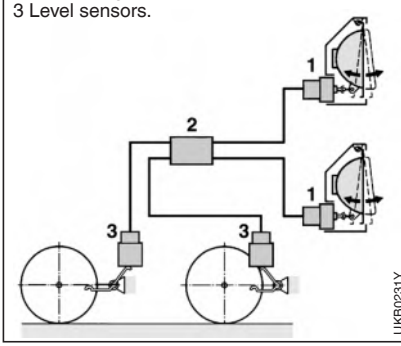
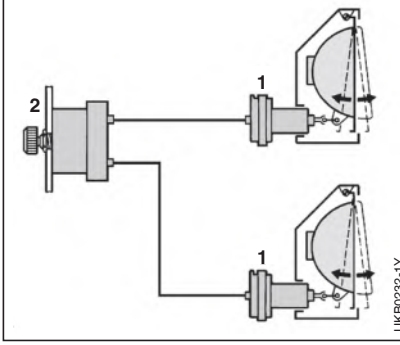


Figure 44: Manual headlamp leveling (principle)
1 Adjustment mechanism,
2 Hand switch.



is projected on a surface vertical to the road surface at the height of the center of the headlamp and above. If, however, the vehicle is subject to extreme variations in attitude due to changes in payload, the headlamps must be adjusted so that the desired aim is achieved.

ECE R48 [12] and EEC 76/756 [24] define the basic setting and the adjustment dimension required on the vehicle. For vehicle categories not covered by these directives, the regulations contained in ECE R48 or ECE R53 [13] shall apply.

Preparations for adjustment
Laden state of vehicle
- Motor vehicles excluding motorcycles: unladen, 75 kg (one person in driver's seat).
- Motorcycles: for EC (as per 93/92/ EEC [25]) unladen, without a person on the driver's seat; for ECE and StVZO (Road Traffic Licensing Regulations) unladen, 75 kg (one person on the driving seat).

Suspension
- Vehicles without self-leveling suspension are rolled for few meters or rocked so that the suspension settles to a level position.
- Vehicles with self-leveling suspension are set and operated to the correct level according to the operating instructions.

Tire pressure
The tire pressure must be adjusted as per the vehicle manufacturer's instructions in accordance with the laden state.

Compulsory function checks on headlamp-leveling control systems
- Systems that operate automatically should be set or operated according to the manufacturer's instructions.
- On vehicles registered before 1/1/1990, a positively locating adjuster position is not required.
- Manually operated systems with two positively locating adjuster positions:
  On vehicles on which the headlamp beam is raised as the vehicle payload is increased, the adjuster must be set to the position at which the beam is at its highest (smallest degree of dip).
  On vehicles on which the headlamp beam is lowered as the vehicle payload is increased, the adjuster must be set to the position at which the beam is at its lowest (smallest degree of dip).

Surface and testing environment
- The vehicle and headlamp tester must be standing on a flat surface (based on ISO 10604 [26]).
- Adjustments and tests should be carried out in an enclosed space where the lighting conditions are not too bright.

Adjusting and testing with a headlamp aiming device
- The headlamp aiming device should be aligned at the specified distance in front of the headlamp under test (except in the case of automatic equipment).
- In the case of headlamp aiming devices which do not run on tracks, vertical alignment relative to the vehicle's longitudinal center axis should be carried out separately for each headlamp. The test should then be carried out without moving the unit sideways again.
- In the case of headlamp aiming devices which run on tracks (or the like), alignment with the vehicle's longitudinal center axis need only be carried out once at the most favorable position (e.g., in a central position in front of the vehicle).
- The specified adjustment dimension for the headlamp concerned should then be set on the headlamp aiming device, and the headlamp adjustment checked and the headlamp adjusted to the correct setting.

Adjusting and testing with a test surface
- The test surface must be vertical to the surface on which the vehicle is standing and at right angles to the vehicle's longitudinal center axis.
- The test surface should be finished in a light color, adjustable vertically and horizontally, and marked with the markings shown in Figure 45.
- The test surface should be positioned at a distance of 10 m in front of the vehicle so that the center mark is aligned with the center of the headlamp under test or adjustment (Figure 46). In the case of lamps with very low beams (e.g.,

fog lamps), a shorter distance can be chosen by calculating the adjustment dimension accordingly.
- Each headlamp must be set individually. Therefore, the other headlamp(s) must be covered over.
- The vertical position of the test surface must be set so that the top cutoff (parallel to the ground) is at the height $h = H - e$. If the surface is not 10 m from the vehicle, the setting e must be converted accordingly.

Figure 45: Test surface for headlamp beam
1 Top cutoff, 2 Center mark, 3 Test surface, 4 Break point.
H Height of headlamp center above standing surface in cm,
h Height of test-surface top cutoff above standing surface in cm,
$e = H - h$ Adjustment dimension.

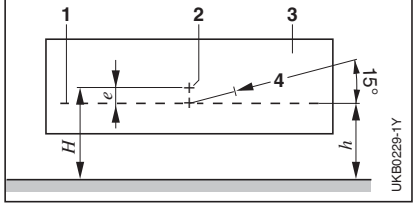
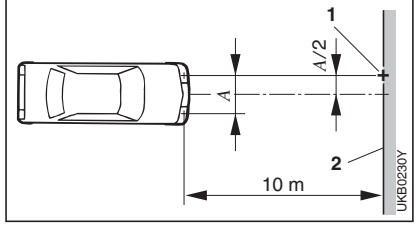


Figure 46: Relative positions of test surface and vehicle longitudinal axis
1 Center mark,
2 Test surface.
A Distance between headlamp centers.



Notes on adjustment
In the case of headlamps with asymmetrical lower beams and fog lamps, the highest position of the light-dark boundary must be on the top cutoff and run as horizontally as possible across the minimum width of the test surface. The lateral adjustment of the headlamps must be such that the light pattern is positioned as symmetrically as possible about the perpendicular line running through the center mark.

In the case of headlamps for asymmetrical lower beam, the light-dark boundary must touch the top cutoff to the left of the center. The intersection between the left section (which should be as horizontal as possible) and the sloping section on the right of the light-dark boundary must be located at the perpendicular line passing through the center mark.

The center of the high beam must be on the center mark.

For headlamps which feature a common adjustment facility for low beam and fog lamps or for high beam, low beam and fog lamps, it is always necessary to use the low-beam headlamp as the basis for adjustment (adjustment dimension e, see Table 5).

Table 5: Headlamp adjustment (excerpt from StVZO (Road Traffic Licensing Regulations))

Vehicle type: Motor veh., > 2 whs. Headlamp position: height above road surface	Adj. dimension "e" Low lamp	Fog beam
Approved to 76/756/EEC or ECE R48 and StVZO, first registered on 1 Jan 1990 or later, < 1200 mm	Adj. dimension on vehicle e.g. ≡D 1.0%	–2.0%
First registered 31 Dec 1989 or earlier, ≤ 1400 mm, and first registered after 31 Dec 1989, > 1200 mm, but ≤ 1400 mm	–1.2%	–2.0%

Headlamp aiming devices
Function
Correct adjustment of motor-vehicle headlamps should ensure the best possible illumination of the roadway by the low beam while minimizing glare for oncoming traffic at the same time. For this purpose, the inclination of the headlamp beams with respect to a level base surface, and their direction to the vertical longitudinal center plane of the vehicle, must satisfy official directives.

Figure 47: Aiming device for headlamps
1 Alignment mirror, 2 Handle,
3 Luxmeter, 4 Refraction mirror,
5 Markings for center of lens.

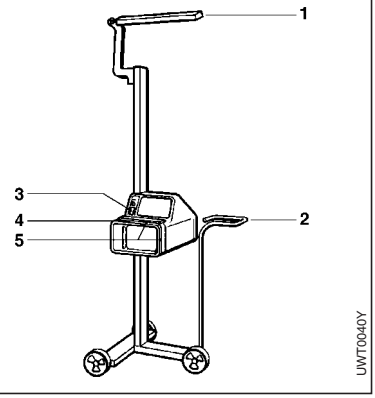
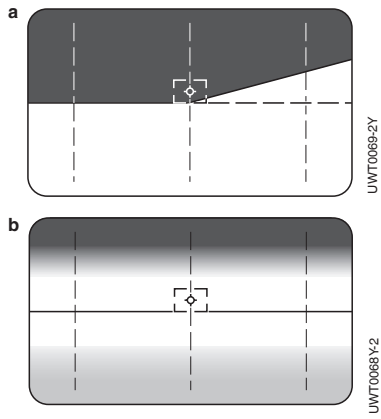


Figure 48: Viewing window in aiming device
a) Top cutoff for light-dark boundary for asymmetrical lower beam,
b) Center mark for middle of upper-beam pattern.



Equipment design
Headlamp aiming devices are portable imaging chambers (Figure 47). They comprise a single lens and an aiming screen located in the focal plane of the lens, and are rigidly connected to it. The aiming screen has markings to facilitate correct headlamp adjustment, and can be viewed by the equipment operator using suitable devices such as windows and adjustable refraction mirrors. The prescribed headlamp adjustment dimension e, i.e., the inclination relative to the centerline of the headlamp in cm at a fixed distance of 10 m, is set by turning a knob to move the aiming screen (Tables 4 and 5).

The aiming device is aligned with the vehicle axis using a sighting device, such as a mirror with an orientation line. It is turned and aligned so that the orientation line uniformly touches two external vehicle-reference marks. The imaging chamber can be moved vertically and clamped at the level of the vehicle headlamp.

Headlamp testing
The headlamp can be tested after the equipment has been correctly positioned at the front of the lens. An image of the light pattern emitted by the headlamp appears on the aiming screen. Some test devices are also equipped with photodiodes and a display to measure the luminous intensity.

On headlamps with asymmetrical lower-beam patterns, the light-dark boundary should touch the horizontal top cutoff; the intersection between the horizontal and sloping sections must be located on the perpendicular line running through the center mark (Figure 48). After adjusting the lower-beam light-dark boundary in accordance with the regulations, the center of the upper beam (assuming that high and low beam are adjusted together) should be located within the rectangular border about the center mark.

Headlamp adjustment, USA

For headlamps compliant with US Federal legislation, visual (vertical only) adjustment as permitted since 5/1/1997 has become increasingly widespread in the US since mid-1997. There is no horizontal aiming here.

Before then, the use of mechanical aiming devices was the most common method of headlamp adjustment in the USA. The headlamp units were equipped with three pads on the lenses – one for each of the three adjustment planes. A calibrating unit is placed against these pads. Aiming is checked using spirit levels. With the VHAD (Vehicle Headlamp Aiming Device) method permitted since 1993, the headlamps are adjusted relative to a fixed vehicle reference axis. This procedure is carried out using a spirit level firmly attached to the headlamp. As a result, the three lens pads were no longer required.

References

[1] StVZO §22a: Design certification for vehicle parts.

[2] StVZO §50: Headlight for high and low beams.

[3] StVZO §52: Additional headlights and lamps.

[4] ECE R3: Uniform provisions concerning the approval of retro-reflecting devices for power-driven vehicles and their trailers.

[5] ECE R4: Uniform provisions concerning the approval of devices for the illumination of rear registration plates of power-driven vehicles and their trailers.

[6] ECE R6: Uniform provisions concerning the approval of direction indicators for power-driven vehicles and their trailers.

[7] ECE R7: Uniform provisions concerning the approval of front and rear position (side) lamps, stop-lamps and end-outline marker lamps for motor vehicles (except motorcycles) and their trailers.

[8] ECE R19: Uniform provisions concerning the approval of motor vehicle front fog lamps.

[9] ECE R23: Uniform provision concerning the approval of reversing lamps for power-driven vehicles and their trailers.

[10] ECE R37: Uniform provisions concerning the approval of filament lamps for use in approved lamp units on power-driven vehicles and their trailers.

[11] ECE R38: Uniform provisions concerning the approval of rear fog lamps for power-driven vehicles and their trailers.

[12] ECE R48: Uniform provisions concerning the approval of vehicles with regard to the installation of lighting and light-signalling devices.

[13] ECE R53: Uniform provisions concerning the approval of category L3 vehicles with regard to the installation of lighting and light-signaling devices.

[14] ECE R65: Uniform provisions concerning the approval of special warning lamps for motor vehicles.
[15] ECE R77: Uniform provisions concerning the approval of parking lamps for power-driven vehicles.
[16] ECE R87: Uniform provisions concerning the approval of daytime running lamps for power-driven vehicles.
[17] ECE R91: Uniform provisions concerning the approval of side-marker lamps for motor vehicles and their trailers.
[18] ECE R98: Uniform provisions concerning the approval of motor vehicle headlamps with gas-discharge light sources.
[19] ECE R99: Uniform provisions concerning the approval of gas-discharge light sources for use in approved gas-discharge lamp units of power-driven vehicles.
[20] ECE R112: Uniform provisions concerning the approval of motor vehicle headlamps emitting an asymmetrical passing beam or a driving beam or both and equipped with filament lamps and/or light-emitting diode (LED) modules.
[21] ECE R113: Uniform provisions concerning the approval of motor vehicle headlamps emitting a symmetrical passing beam or a driving beam or both and equipped with filament lamps.
[22] ECE R119: Uniform provisions concerning the approval of cornering lamps for power-driven vehicles.
[23] ECE R123: Uniform provisions concerning the approval of adaptive front-lighting systems (AFS) for motor vehicles.
[24] 76/756/EEC: Council Directive of 27 July 1976 on the approximation of the laws of the Member States relating to the installation of lighting and light-signaling devices for motor vehicles and their trailers.
[25] 93/92/EEC: Council Directive of 29 October 1993 on the installation of lighting and light-signalling devices on two or three-wheeled motor vehicles.
[26] ISO 10604: Road vehicles; measurement equipment for orientation of headlamp luminous beams.
[27] FMVSS 108: Lamps, reflective devices, and associated equipment.
[28] SAE J186: Supplemental High Mounted Stop and Rear Turn Signal Lamps for Use on Vehicles Less than 2032 mm in Overall Width.
[29] SAE J222: Parking Lamps (Front Position Lamps).
[30] SAE J585: Tail Lamps (Rear Position Light) for Use on Motor Vehicles Less than 2032 mm in Overall Width.
[31] SAE J586: Stop Lamps for Use on Motor Vehicles Less Than 2032 mm in Overall Width.
[32] SAE J587: License Plate Illumination Devices (Rear Registration Plate Illumination Devices).
[33] SAE J588: Turn Signal Lamps for Use on Motor Vehicles Less Than 2032 mm in Overall Width.
[34] SAE J592e: Clearance, Side Marker and Identification Lamps.
[35] SAE J593c: Back-up Lamps.
[36] SAE J594: Reflex Reflectors.
[37] SAE J852: Front Cornering Lamps for Use on Motor Vehicles.
[38] D. Meschede: Gerthsen Physik. 24th Edition, Springer-Verlag, 2010.
[39] R. Baer: Beleuchtungstechnik – Grundlagen. 3rd Edition, Huss-Medien-GmbH, Verlag Technik, 2006.

Automotive glazing

The material properties of glass

Basic components

Window panes for automotive use are made of silica glass. The basic chemical constituents and their proportions are as follows:

- 70 to 72 % silicic acid (SiO_2) as the basic component of glass,
- Approx. 14 % sodium oxide (Na_2O) as flux,
- Approx. 10 % calcium oxide (CaO) as stabilizer.

These substances are mixed in the form of quartz sand, soda ash, and limestone. Other oxides such as magnesium and aluminum oxide are added to the mixture in proportions of up to 5 %. These additives improve the physical and chemical properties of glass.

Manufacturing flat glass

The glass panes are made out of the basic product, flat glass. Flat glass that is cast using the float-glass process is used. This process involves melting the mixture at a temperature of 1,560 °C. The melt then passes through a refining zone at 1,500 to 1,100 °C, and is then floated on a float bath of molten tin. The molten glass is heated from above (smoothing of the surface by fire-finishing). The flat surface of the molten tin creates flat glass with flat parallel surfaces of a very high quality (tin bath underneath, fire-finishing on top). The glass is cooled to 600 °C before it is lifted out of the float bath into the cooling section. After a further period of slow, non-stress cooling, the glass is cut into sheets measuring 6.10 × 3.20 m^2.

Tin is suitable for the float-glass process because it is the only metal that does not produce any vapor pressure at 1,000 °C and is liquid at 600 °C.

Table 1: Material properties and physical data of glass and finished windshields and windows

Property	Dimension	TSG	LSG
Density	kg/m^3	2,500	2,500
Hardness	Mohs	5 to 6	5 to 6
Resistance to pressure	MN/m^2	700 to 900	700 to 900
E module	MN/m^2	68,000	70,000
Flexural strength			
Before pre-tension	MN/m^2	30 [2]	30 [1]
After pre-tension	MN/m^2	50 [2]	
Specific heat	$kJ/kg \cdot K$	0.75 to 0.84	0.75 to 0.84
Thermal conduction coefficient	$W/m \cdot K$	0.70 to 0.87	0.70 to 0.87
Coefficient of thermal expansion	K^{-1}	$9{,}0 \cdot 10^{-6}$	$9{,}0 \cdot 10^{-6}$
Relative permittivity		7 to 8	7 to 8
Light transmittance (DIN 52306) [1] clear [3]	%	≈ 90	≈ 90 [1]
Refractive index [3]		1.52	1,52 [1]
Deviation angle of wedge [3]	Arc minute	< 1.0 flat < 1.5 curved	≤ 1.0 flat [1] ≤ 1.5 curved [1]
Dioptric divergence DIN 52305 [2] [3]	Diopters	< 0.03	≤ 0.03 [1]
Thermal stability	°C	200	90 [1] (max. 30 min)
Resistance to temperature shocks	K	200	

[1] Properties of finished laminated safety glass (LSG). In calculating the permissible bending stress, the coupling effect of PVB film is to be disregarded.
[2] Calculated values; these values already contain the necessary safety margins.
[3] Figures for optical properties depend very much on the type of window.

Automotive windshield and window glass

Glass used for automotive glazing is of two types:
– Single-pane toughened safety glass (TSG), which is chiefly used for side-window, rear-window and sunroof glazing.
– Laminated safety glass (LSG), which is used primarily for windshields and rear windows, but also for sunroofs. LSG is also increasingly being fitted in vehicle side and rear windows.

The materials from which TSG and LSG panes are made are the basic glass types:
– Transparent float glass: This glass offers the best possible light transmittance.
– Tinted float glass: This glass has a homogeneous green or gray tint within the material; the tint blocks heat from the sun.
– Coated float glass: The glass is coated on one side with noble-metal and metal oxides; the coating reduces heat and UV radiation entering the vehicle, thus providing thermal insulation.

TSG panes
TSG panes differ from LSG panes as they have greater mechanical and thermal strength, and their breaking and shattering behavior is different. They pass through a toughening process which greatly prestresses the surface of the glass. In case of breakage, these panes shatter into many small blunt-edged pieces (to eliminate the risk of injury).

Post-processing (by grinding or drilling) of TSG panes is not possible. The standard thicknesses are 3, 4 and 5 mm.

LSG panes
An LSG pane is made of a crack-proof, flexible plastic intermediate layer of polyvinyl butyral (PVB) bonded between two sheets of glass. When subjected to impact or shock, the glass splits into web-like crack patterns. The plastic intermediate layer holds the broken pieces of glass together (to eliminate the risk of injury). The laminate retains its integrity and transparency when the glass is shattered.

The standard thicknesses for LSG panes are 4 to 5 mm.

Optical properties
The requirements for the optical quality of automotive glazing are as follows:
– unimpeded vision,
– flawless vision,
– undistorted vision.

Achievement of optimum optical quality has to be balanced against structural requirements and the vehicle-body design, taking account of such factors as
– glazing with large surface areas,
– glazing that is fitted laying flat,
– cylindrical or spherical panes,
– panes with a high degree of curvature.

Possible quality impairments arise from:
– optical deflection,
– optical distortion,
– double imaging.

Optical deflection increases with:
– increasing obliqueness of the angle of incidence, i.e., increasing slope of the window,
– increasing pane thickness,
– reducing radius of curvature (increasing degree of bend),
– increasing divergence from perfect surface parallelism of the original glass sheet.

Green- or gray-tinted glass is used as heat absorption glass because it blocks the transmission of infrared rays (heat radiation) more strongly than shorter wavelengths. On the other hand, it also reduces transmittance within the visual spectrum. The PVB film in LSG absorbs light in the ultraviolet range.

The optical properties of TSG and LSG panes are roughly the same because the optical properties of the intermediate plastic layer in LSG are very similar to those of glass in the visible spectrum.

Functional design glazing

The demands placed on glazing are continually increasing. The flat panes used in the past simply served to protect occupants from wind and weather. Now, automotive glazing performs a wide variety of functions.

Tinted glazing
This is made from glass which is tinted within the material, and reduces the direct penetration of solar radiation into the vehicle interior.

The reduction in the transmittance of solar energy occurs chiefly in the long-wavelength spectrum (infrared, Figure 1), and this mainly lowers the transmittance of energy, thus reducing the heat transmitted to the vehicle interior. The degree to which light transmittance within the visible spectrum is affected depends on the degree of color tinting and the thickness of the glass.

For windshields, light transmittance must be at least 70%. Strongly tinted glass with a light transmittance of less than 70% can be used in windows from the B-pillar to the rear, if the vehicle has two exterior mirrors. For sunroof glazing, tinted glass with a significantly lower light transmittance is used.

Coated glazing
This type of glazing uses glass with a metal or metal-oxide coating. Depending on the production process, the coating may be carried out before or after the sheet of glass is bent and prestressed. The coating is applied to the inner glass surface of an LSG pane.

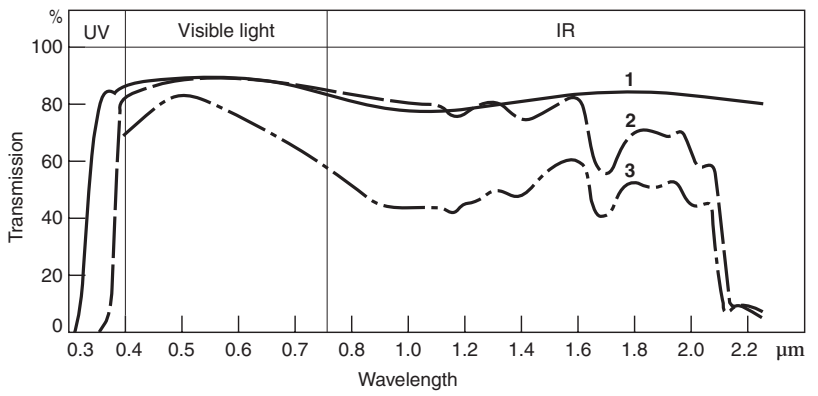
This type of coated glazing has a light transmittance of less than 70%. Consequently, it can be fitted from the B-pillar to the rear, if the vehicle has two exterior mirrors.

Coated glass can also be used for sunroof glazing. Another type of sunroof glazing uses pyrolytically coated windows which are post-processed after the coating has been applied.

Coated glazing reduces direct solar radiation into the vehicle interior and absorbs solar energy in the infrared and ultraviolet ranges.

Windshields with sunshield coating
A coating is applied to the inner surface of the outer or inner layer of glass in a laminated glass pane. The coating is a multilayer interference system with silver as its base layer. As the coating is on the inside of the laminate, it is permanently protected against corrosion and scratching.

Figure 1: Light transmittance of automotive windshield and glass
1 Float glass and TSG windows, thickness 4 mm, non-tinted,
2 LSG windows, overall thickness 5.5 mm, non-tinted,
3 LSG windows, overall thickness 5.5 mm, green.
UV Ultraviolet,
IR Infrared.



The purpose of the coating is to reduce the transmittance of solar energy by more than 50 %. thus reducing the heat transmitted to the vehicle interior. Transmittance is reduced primarily in the infrared range so that visible light transmittance is altered negligibly. The reduction is achieved largely by reflection, so that secondary reflection into the interior is low. The light transmittance of this glass in the UV range is very low, i.e., less than 1 %.

Sunroof glazing made from laminated safety glass

The bent laminated safety glass consists of two tinted-glass layers which are thermally partially prestressed in order to increase mechanical strength. They are bonded to either side of a highly crack-resistant and specially tinted film. The overall thickness depends on the surface area of the glass, and the overall design of the sunroof.

Absorption that takes place mainly in the infrared range guarantees minimum heat penetration. The coating also provides low light transmittance and complete filtering of UV light.

Automotive insulation glass

Automotive insulation glass is made of two flat or curved sheets of single-pane toughened safety glass (3 mm) separated by an air gap (3 mm). This type of insulation glass reduces heating of the vehicle interior, particularly in combination with a tinted coating. Transmittance is reduced primarily in the infrared range so that visible light transmittance is altered negligibly.

Insulation glass also provides better thermal insulation in the winter and improves sound insulation.

Insulation glass is now only used on commercial vehicles such as buses, trains and airplanes.

Heated laminated safety glass

Heated laminated safety glass can be used for windshields or rear windows. They prevent icing and misting of the glass pane, even in extreme winter temperatures, and ensure clear visibility.

Heated laminated safety glass is made up of two more sheets of glass with a PVB film bonded between them. Inside the PVB film, there are heater filaments which may be less than 20 µm thick, depending on the heat output required. The heater filaments may be laid in a waveform pattern or in straight lines. They may also run vertically or horizontally. The heated area may cover the whole of the pane or may be divided into separate zones with varying heat outputs. In this way, the windshield wipers can be prevented from freezing to the windshield in subzero temperatures.

Heated windows can also be created by layering. In contrast to filament heating, these layers have the advantage of being invisible and reflecting infrared radiation.

Automotive antenna glass
This type of glass has antenna wire embedded in it. In the case of windows made of single-pane toughened safety glass (roof and side windows), the antenna is printed on the glass and is located – almost invisibly – on the inner surface of the window. With laminated safety glass windows (windshields), the antenna wiring system is embedded or printed onto the inner film.

Acoustic glazing
An acoustic windshield is composed, like every windshield, of two individual sheets of glass – an outer and an inner sheet which are firmly connected to each other by way of a sturdy plastic film of polyvinyl butyral (PVB). In the case of acoustic glazing, however, the conventional PVB film is replaced by acoustic PVB which has a highly damping acoustic core located in the middle between two layers of standard PVB. The two outer layers of film of conventional plastic thus safeguard the mechanical properties, while the core layer of damping material absorbs vibrations.

Acoustic windshields reduce the transmission of low-frequency engine vibrations into the vehicle interior. Engine-vibration noises are absorbed by the windshield, these booming noises thus being reduced at low frequencies by up to 5 dB. Side windows can also be fitted with acoustic glazing in the meantime.

Water-repellent glazing
A coating on the glass causes tiny pearls of water to form on the glass surface when it rains. These are blown off the glass by the air stream.

All glass surfaces can be fitted with this function. After several years of use, the effectiveness of water-repellent glazing can be fully restored with a regenerating kit.

Panorama roofs
There is a wide selection of panorama roofs available, ranging from the closed, large-surface glass roof, through the twin- or multi-panel roof, to the lamellar sunroof.

The following glass technologies can be used for panorama roofs:
– conventional laminated safety glass, 5 mm thick,
– partially prestressed laminated safety glass, 5 mm thick.

Switchable glazing
There are various technologies for influencing the transparency of a pane “at the touch of a button”. One of the common methods is based on liquid crystal which are embedded in plastic foil. When a voltage is applied, all the crystals are orientated in one direction with the result that the pane is switched from opaque to transparent. The switching duration is under one second.

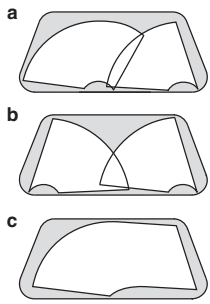
HUD glazing
A head-up display projects a variety of information onto the windshield (e.g., driving speed, navigation, road-sign detection) while the vehicle is in motion. Unlike a conventional windshield, an HUD windshield uses PVB film which increases minimally in thickness from the engine edge to the roof edge. This produces a sharp HUD image of high optical quality. The optimum configuration between projector and windshield is being established as part of elaborate simulation and development work.

References
[1] DIN 52306: Ball drop test on safety glass for vehicle glazing.
[2] DIN 52305: Determination of the optical deviation and refractive power of safety glass for vehicle glazing.

Windshield, rear-window and headlamp cleaning systems

The windshield and rear-window cleaning systems is to provide the driver with sufficient visibility in the motor vehicle at all times. The following different types of system are used:
– front wiper systems,
– rear wiper systems,
– windshield and rear-window washer systems in combination with wiper systems,
– headlamp cleaning systems.

Figure 1: Front wiper systems for fulfilling required passenger-car areas of vision
a) Tandem system,
b) Opposed-pattern system,
c) Wiper-blade-controlled single-arm system.



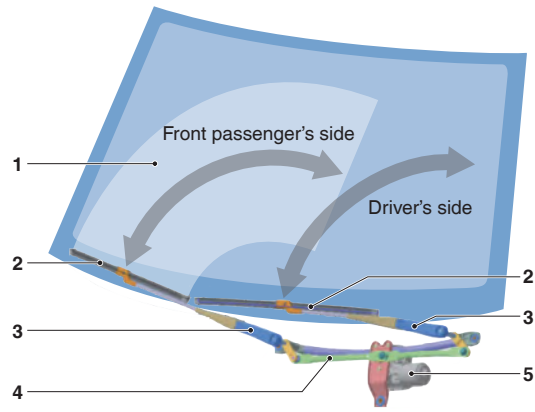
Front wiper systems

Functions and requirements
The function of front wiper systems is to remove water, snow and dirt (mineral, organic or biological) from the windshield. Boundary conditions are among others:
– operational at high temperature (+80 °C) and low temperature (–40 °C),
– corrosion resistance against acids, bases, salts, ozone,
– endurance strength under load (e.g., due to high snow load),
– pedestrian impact protection,
– cleaning effect at high driving speeds,
– low operating noise.

Legislators have translated the demand for adequate visibility into standardized areas of vision on the windshield (e.g., for Europe [1]). These are subdivided into several areas and must be cleaned by the wiper system to fixed percentage levels. The most important wiper systems for passenger-car windshields which satisfy these requirements are shown in Figure 1.

Wiper systems for commercial vehicles are similar to those for passenger cars, but are subject to different requirements especially with regard to driving speed and windshield shape.

Figure 2: Front wiper system
1 Windshield,
2 Wiper blade,
3 Wiper arm,
4 Wiper linkage,
5 Wiper drive.



Operating principle

Essentially a front wiper system consists of a wiper drive with or without linkage, the wiper arms, and the wiper blades (Figure 2).

The wiper drive moves the wiper arms at a specific angle over the windshield so that the driver and front passenger have a clear view. A specially shaped rubber wiper lip provides for an optimal wiper result.

Front wiper drive

Motor-gear unit

Front wiper drives essentially consist of an electric motor with a reduction gear. Mainly permanent-field DC motors with brushgear (mechanical commutation) are used as electric motors and helical-gear mechanisms are used as gears (Figure 3).

Brushless permanent-field DC motors with electronic commutation are for cost reasons still in the minority and only appear occasionally on the market.

Motor design

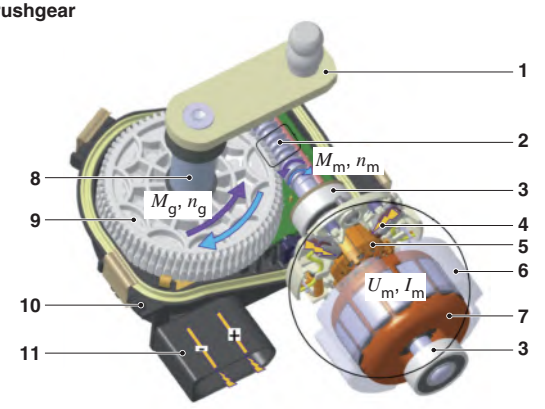
The motors are for cost, weight and space reasons high-speed in design. Adaptation to the speed and torque requirements of the wiper system is performed via the helical-gear mechanism.

The primary function of wiper drives is to guarantee visibility under anticipated conditions, i.e., among others sufficiently frequent windshield wiping. For front wiper systems this usually means around 40 timers per minute (setting 1) and under more extreme conditions around 60 times (setting 2). This is stipulated in Europe [1] with the demand for a wiping frequency of at least 10 and at most 55 wiping cycles per minute and a second wiping frequency of at least 45 complete wiping cycles per minute. The difference between the highest and lower wiping frequency must be at least 15 wiping cycles per minute.

The classification by power output often used for electrical machines is not suitable for drives for wiper systems. Wiper systems are normally operated on a wet windshield; in this case, minimal torque is demanded of the motor at high rotational speeds. On the other hand, in the case of high-load states such as when the wiper blade is frozen, a high torque is required at low rotational speed. In both cases, the power output (product of torque and rotational speed) is low. Because the drive variable is dependent on the required torque at the high load point, this torque is used as the classification feature.

Figure 3: Wiper motor with brushgear and helical-gear mechanism

1 Motor crank,
2 Worm,
3 Bearing,
4 Carbon brush,
5 Commutator,
6 Magnet,
7 Armature,
8 Output shaft,
9 Helical gear,
10 Gear-mechanism cover,
11 Electrical connection.
U Voltage,
I Current,
M Torque,
n Rotational speed,
m Motor,
g Gear.



Rotary drives
Rotary drives (Figure 4) are characterized by the fact that the output shaft of the wiper drive rotates permanently in the same direction (360°). The actual wiping motion on the windshield is determined by the wiper linkage and its kinematic configuration.

The speed/torque equation and the starting torque derived from it are governed by the following relationships:

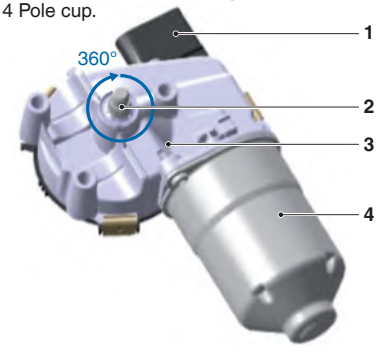
$$n = -2\pi \frac{R_A}{c^2 \Phi^2} M + \frac{U_{KL}}{c\Phi}\ , \qquad (1)$$

$$M_A = \frac{1}{2\pi} \frac{U_{KL}}{R_A} c\Phi\ . \qquad (2)$$

$c = zp/a$ Motor constant,
U_{Kl} Terminal voltage,
Φ Interlinked winding flux,
R_A Armature resistance,
M Torque,
M_A Starting torque for $n = 0$,
z Effective total number of conductors,
$2a$ Number of parallel armature-winding paths,
p Number of pole pairs.

The two speed settings are obtained in such systems by having carbon brushes to ground (–) and to battery voltage (+) for setting 1 and switching to a third carbon brush (Figure 5), which is also connected to battery voltage (+) and delivers setting 2.

Figure 4: Rotary wiper drive
1 Gear-mechanism cover with electrical connection, 2 Output shaft,
3 Gear-mechanism housing,
4 Pole cup.



In response to the effect of the third brush there are situated under a magnet conductors which bring about a braking torque. When this braking torque is subtracted

Figure 5: Position of third carbon brush
1 Magnet, 2 Carbon brushes for setting 1,
3 Third brush for setting 2.

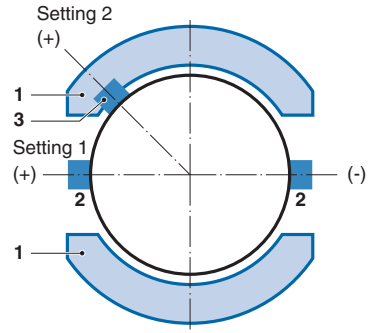
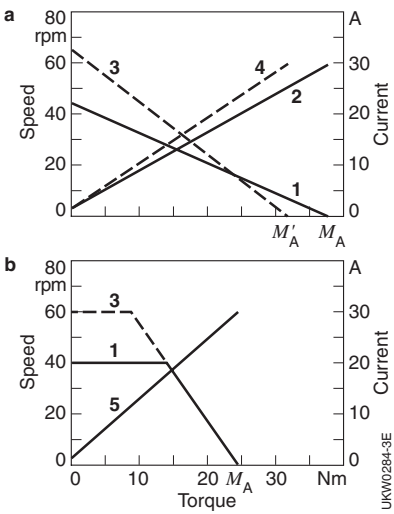


Figure 6: Speed/torque curves for wiper drives
a) Rotary wiper drive,
b) Reversing wiper drive.
1 Speed setting 1,
2 Current setting 1,
3 Speed setting 2,
4 Current setting 2,
5 Current, reversing mode.
M_A Starting torque.



from the useful torque of the conductors under the other magnet, the third brush, which concerns the number of conductors, acts is if to reduce the number of conductors for a constant wire cross-section, i.e., the speed/torque curve should rotate about the starting point M_A (Figure 6a). When the armature resistance is being determined, however, these conductors cannot be ignored, producing an armature resistance that is greater that the torque effect from the perspective of the number of conductors. According to equation (2), therefore, the starting torque will decrease and a center of rotation of the curve will appear which is in the rear part of speed/torque curve for M_A^* (Figure 6a).

The wiper drives have a mechanical sensor system to ensure that the wiper blades adopt the correct parking position. Because the position of the output shaft of the rotary drive has a 1:1 correlation with the wiping angle, this function can be guaranteed with a reference position at the output. Mainly sliders in the gear-mechanism cover with a contact disk in the helical gear are used as the parking-position system (Figure 7).

High-load and locking protection is provided in rotary drives conventionally by means of a thermoswitch.

Reversing drives

In the case of reversing drives the output shaft oscillates at a defined angle, usually less than 180° (Figure 8). Output to the wiper shafts is provided, as is also the case with rotary wiper drives, via a wiper linkage or in direct drives directly to the wiper shaft.

The gear-mechanism cover incorporates electronics with software. These induce, by reversing the polarity of the terminal voltage, a reversal of the electric-motor direction and with it a reversal of direction of the wiper blades at the corresponding turning points. Furthermore, pulse-width modulation of the applied terminal voltage is used to generate different motor speeds and thus controllable wiper speeds or wiper frequencies (Figure 6b). Therefore, reversing drives with a DC motor with brushgear and mechanical commutation, unlike conventional rotary drives, have only two carbon brushes.

The position and speed of the output is relevant to controlling the drive. For example, these variables are detected via Hall sensors integrated in the electronics and sensor magnets in the helical gear and processed with corresponding software. The signals from this sensor system are also used to detect the parking position (Figure 9) and spare the mechanical parking-position system of conventional rotary drives.

Figure 7: Mechanical parking-position system

1 Gear-mechanism cover with slider,
2 Contact disk
3 Helical gear.

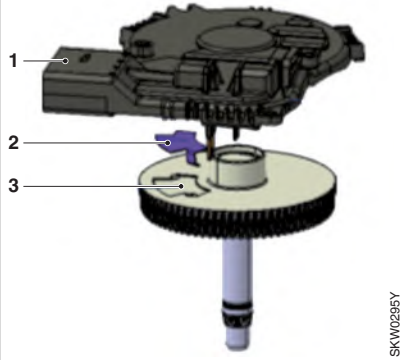
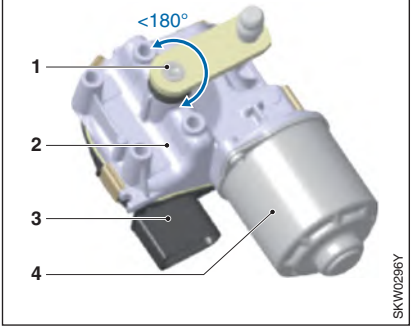


Figure 8: Reversing wiper drive

1 Output shaft,
2 Gear-mechanism housing,
1 Gear-mechanism cover with electrical connection,
4 Pole cup.



High-load and blocking protection is likewise guaranteed here by the integrated electronics.

Wiper linkage
The wiper linkage is the connection between the drive and one or more wiper arms. It transfers the movement of the wiper drive via the motor crank with ball pivot to the articulated rods and via these to the ball pivot of the wiper bearing (Figure 10).

The wiper arm is connected by means of a collar nut for example to the wiper-bearing shaft and the wiper drive is connected by means of a plate and corresponding screws in the gear-mechanism housing to a shaped tube. The shaped tube simplifies installation in the vehicle and limits the wiping-angle tolerances.

The following different design variations are available.

Rotary drive
The drive runs in a circle (360°) and the linkage transfers the circular motion into the oscillating motion of the wiper arms (Figure 11a). The requested wiping angle determines the kinematic configuration of the linkage, i.e., the angle between the upper stop of the wiper blade near the A-pillar (upper turning point) and the lower stop near the bottom edge of the windshield (lower turning point).

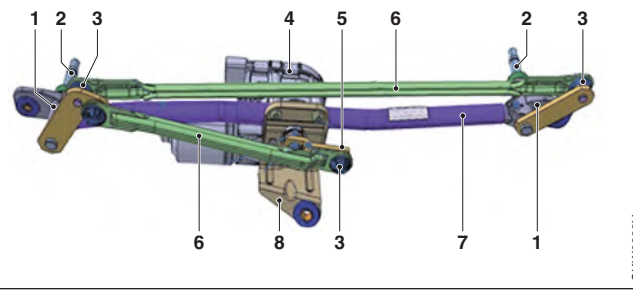
Reversing technology
The electronically controlled drive rotates less than a half rotation (<180°). The linkage transfers the force from the drive to the wiper arms (Figure 11b). The advantage of such systems over the rotary drive lies in the fact that the space required for the linkage kinematics is roughly halved.

In direct drives the wiper arm is directly connected to the output shaft. Here there are systems with reduced linkage complexity (Figure 11c), omitted shaped tube (Figure 11d), and systems entirely without linkage (Figure 11e).

Figure 9: Electronic parking-position system
1 Gear-mechanism cover with electronics and sensor,
2 Sensor magnet,
3 Helical gear.



Figure 10: Basic wiper linkage
1 Wiper bearing,
2 Wiper-bearing shaft,
3 Ball pivot,
4 Wiper drive,
5 Motor crank,
6 Articulated rod,
7 Shaped tube,
8 Plate.

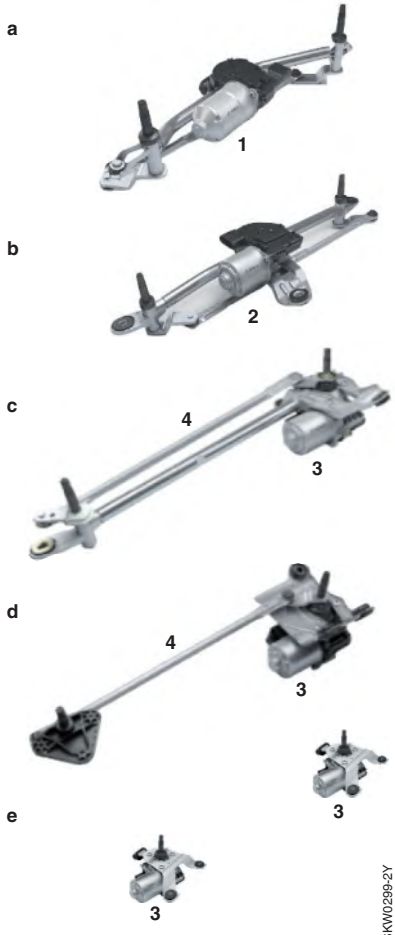


Omitting the shaped tube cuts down on the number of wiper-linkage parts, but increases the amount of fitting work by the vehicle manufacturer in that more parts must be connected to the vehicle. Furthermore, this increases the demands placed on the rigidity of the vehicle body and necessitates increased precision when installing the components in order to ensure wiping-angle precision.

When the wiper linkage is omitted, the drive is directly screwed to the body by means of a plate.

Additional functions, such as, for example, an extended parking position, an overload-protection function (e.g., for snow), an infinitely variable wiping speed, and a consistently large wipe pattern over varying operating conditions (e.g., driving speed), can be implemented with electronically controlled wiper systems.

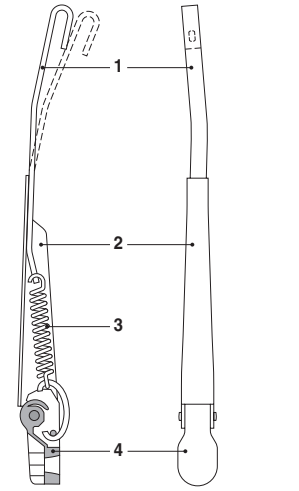
Figure 11: Linkage configurations
a) Rotary drive with linkage,
b) Reversing drive with linkage,
c) Reversing direct drive with linkage,
d) Reversing direct drive without shaped tube,
e) Two reversing direct drives without linkage.
1 Rotary drive,
2 Reversing drive,
3 Reversing direct drive,
4 Wiper linkage.



Wiper arms

The wiper arm is the connecting element between the wiper linkage and the wiper blade. It is screwed to the tapered wiper-bearing shaft by its mounting end, which is usually made of diecast aluminum or steel sheet (Figure 12). The other end is generally a steel strip and holds the wiper blade. A distinction is made here

Figure 12: Wiper arm (side and top views)
1 Steel strip with hook-type fastening,
2 Joint section,
3 Tension spring,
4 Attachment part with taper for mounting on wiper-bearing shaft.



between hook-type, side and front fastenings (Figure 13).

To further improve the visibility conditions the washer fluid of the windshield washer system is specifically directed to where it is needed, i.e., directly in front of the wiper blade. Nozzle bodies (spray nozzles) are integrated in the wiper arm (Figure 14) and spray holes integrated in the wiper blade (Figure 15) for this purpose. This improves the cleaning effect, widens the field of vision, and facilitates a clear view. Unlike conventional systems, there is no wide-area spray mist. This additionally improves the view event at higher driving speeds. Depending on the available equipment specification in the vehicle, these systems come heated or unheated.

Figure 13: Fastening the wiper blade to the wiper arm
a) Hook-type fastening,
b) Side fastening,
c) Front fastening.

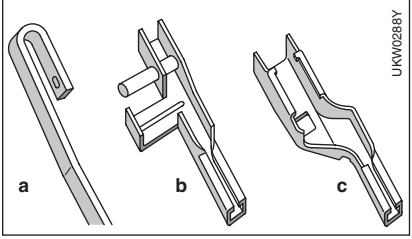


Figure 14: Wiper arm with nozzle body
1 Wiper blade,
2 Nozzle body,
3 Wiper arm,
4 Electrohydraulic connection.

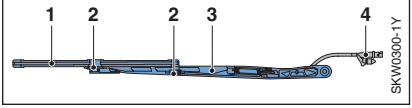
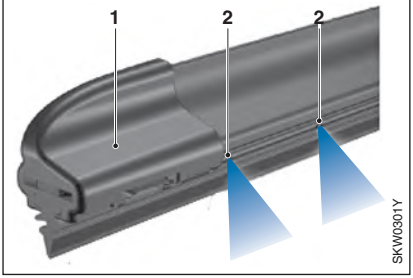


Figure 15: Wiper blade with spray holes
1 Wiper blade,
2 Spray holes.



The wiper bearing is screwed to the body. The position of the wiper arm in relation to the windshield is established by way of the position of the wiper bearing relative to the windshield. The field of vision is established in combination with the wiper-blade length and the wiping angle.

The position of the wiper blades is defined by the angular position of the wiper bearings to the windshield and by applying additional torsion to the wiper arms. The aim is to incline the wiper arms laterally at their reversal points towards the wipe-angle bisector. This helps the wiper-blade elements to swivel into their new working position. This, in turn, reduces wiper-blade wear and the reversing noise.

Wiper blades

The wiper blade holds the wiper-blade element and guides it over the windshield. Essentially there are two different types: conventional wiper blades (Figure 16) and jointless wiper blades (Figure 17). Wiper blades are used in lengths of 260 to 800 mm. Their dimensions for mounting (e.g., hook-type or snap-in fastening) are standardized.

Conventional wiper blade

In conventional wiper blades the contact pressure is distributed by claw brackets (Figure 16). Low-wear operation is achieved by compensating the play in the mountings and linkages. The tops of the center brackets are perforated to prevent blade liftoff at high speeds. In special cases, wind deflectors are integrated into the wiper arms or blades to press the blades against the windshield.

Jointless wiper blade
The flat, jointless wiper blade (aero-wiper blade) represents the current development status in wiper-blade design (Figure 17).

The contact pressure over the wiper-blade element is no longer distributed by the claws of the wiper bracket, but by two preshaped spring strips (sprung rails) specially adapted to the shape of the windshield. They press the wiper-element lip more evenly against the windshield. This reduces wear on the wiper-element lip and increases wipe quality. In addition, the elimination of the bracket system means that there is no linkage wear, the overall height of the wiper system is substantially reduced, the weight is lower, and the wiper is quieter (there is also less wind noise).

The top edge of the wiper blade is in the shape of a spoiler (air deflector) and allows the blade to be used even at very high speeds without any modification. The flexible material of the spoiler is also much less likely to cause injury to pedestrians in the event of an accident (pedestrian impact protection).

An adapted, simplified connection to the wiper arm ensures a reliable attachment of the wiper blade when the wipers are in operation, and also allow easy replacement when required.

Wiper-blade element
A crucial element of the wiper system is the wiper-blade element. It is retained by the claws of the bracket system (Figure 16) or supported by spring strips (Figure 17). The edge of the rubber lip touches the windshield over a width of 0.01 to 0.015 mm. When moving across the windshield, the wiper-blade element must overcome coefficients of dry friction of 0.7 to 1.7 (depending on air humidity), and coefficients of wet friction of 0.1 to 0.5

Figure 16: Conventional wiper blade
a) Wiper blade under load,
b) At zero load,
c) Cross-section.
1 Wiper-blade element, 2 Claw bracket,
3 Joint, 4 Center bracket, 5 Adapter,
6 Spring strip.

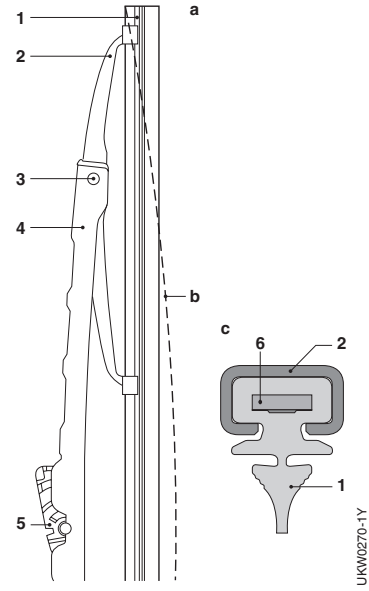
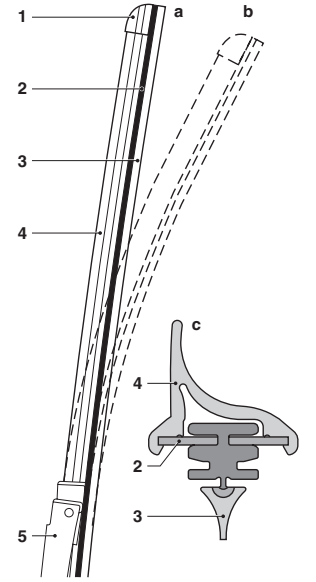


Figure 17: Jointless wiper blade (aero-wiper blade)
a) Wiper blade under load,
b) At zero load,
c) Cross-section.
1 End clip, 2 Spring strip, 3 Wiper-blade element, 4 Spoiler, 5 Wiper arm.



(depending on frictional velocity). The correct combination of wiper-element profile and rubber properties must be chosen so that the wiper lip can cover the complete wipe pattern on the windshield surface at an angle of approx. 45° (Figure 18).

The two-component element (twin wiper), made of synthetic rubber, consists of a specially hardened, abrasion-resistant lip which merges into an extra-soft spine. The soft spine ensures that the wiper element has optimum reversing characteristics and wipes smoothly.

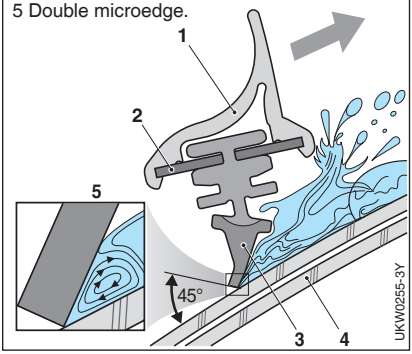
Rain sensor

The rain sensor (see Sensors) detects how heavy the rain is, and converts this data into an appropriate signal that is sent to the wiper drive. This is automatically switched on and, depending on the requirement, operated in the intermittent setting or in setting 1 or setting 2.

The full potential of the rain sensor is fully realized with an electronically controlled wiper system, whose wiping speed can be infinitely re-adjusted to the rain intensity.

Figure 18: Wiper-blade element in working position.

1 Spoiler,
2 Spring strip,
3 Wiper-blade element,
4 Windshield,
5 Double microedge.



Rear wiper systems

Function and requirements

Rear wiper systems are used when the rear window, on account of its sloping angle or the body shape, is prone to getting very dirty and rear visibility is compromised. The principle of rear-window cleaning is generally the same as that for windshield cleaning systems.

A rear-window cleaning system is subject to much lower demands than those of a front wiper system. The rear wiper system is therefore frequently operated in intermittent mode and its field of vision is not subject to any statutory requirements. The wiping angle typically ranges between 60° and 180° (Figure 19).

Figure 19: Rear-window wipe patterns.

The gray areas represent impaired visibility when vehicles are overtaking.

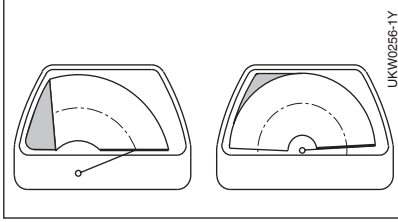
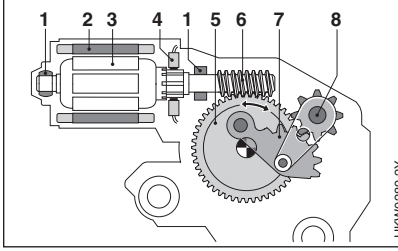


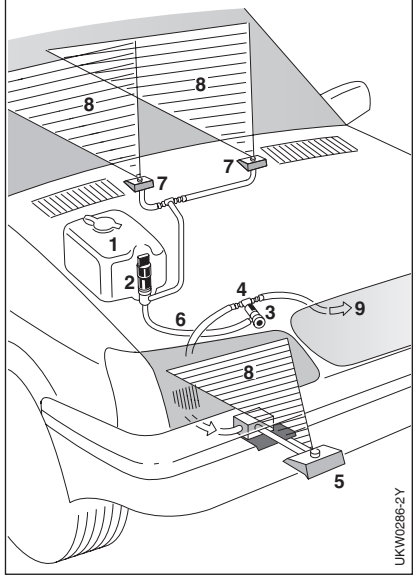
Figure 20: Rear wiper drive with oscillating mechanism.

1 Bearing, 2 Magnet, 3 Armature,
4 Commutator with carbon brushes,
5 Helical gear, 6 Worm, 7 Oscillating mechanism, 8 Output shaft.



Rear wiper drives
Rear wiper drives are essentially no different from front wiper drives. A rear-window wiper system is usually driven by an electric motor with integrated oscillating mechanism, which assumes the function of the linkage in front wiper systems and performs the oscillating movement of the output shaft (Figure 20). The wiper arm is directly attached to the output shaft of the wiper drive.

Figure 21: Windshield washer system and high-pressure washer system for headlamps.
1 Washer-fluid reservoir, 2 Pump,
3 Non-return valve, 4 T-joint,
5 Nozzle holder for headlamp cleaning,
6 Tube,
7 Nozzles of windshield washer system,
8 Spray area,
9 Tube to left headlamp.



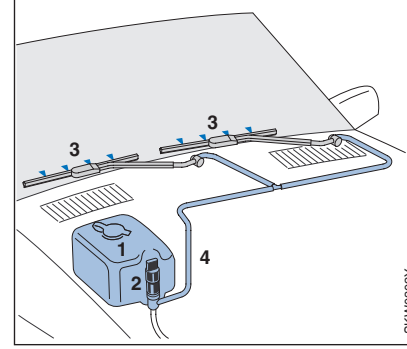
Windshield and rear-window washer systems

Design
To ensure good cleaning, it is imperative that the wiper system is backed by a washer system. Electrically driven centrifugal pumps are used to direct water mixed with detergent additive from the washer-fluid reservoir through nozzles in a pointed spray or through sprinkler nozzles as water mist onto the windshield (Figure 21).

To spray the windshield with water, the nozzles are positioned on the engine hood or in line with current development trends directly on the wiper arm, or the spray holes are integrated in the wiper blade (Figure 22). The capacity of the washer-fluid reservoir is usually 1.5 to 2 *l*. If the same washer-fluid reservoir is also used to supply the headlamp washers, a capacity of up to 7 *l* may be required. A separate reservoir may be provided for the rear-window cleaning system.

Electronic control
The washer system is often linked to the cleaning system by means of an electronic control system so that water is sprayed onto the rear window or windshield for as long as a pushbutton remains pressed. The wiper system then continues to operate for several additional cycles after the pushbutton is released.

Figure 22: Windshield washer system with nozzles integrated in the wiper arm.
1 Washer-fluid reservoir, 2 Pump,
3 Wiper blade with spray holes integrated in the wiper blade,
4 Tube.



Headlamp cleaning systems

Pure washer systems have become established for cleaning headlamps. The advantages of the headlamp washer system over the previously used wiper systems lie in its simple design and in the fact that it is easier to adapt to the vehicle styling concept.

A headlamp washer system is prescribed by law in Germany for xenon headlamps to prevent oncoming traffic from being dazzled by light diffusion.

Design

High-pressure washer systems (Figure 21) consist of the washer-fluid reservoir (the washer fluid required is taken from the reservoir for the windshield washer system), the pump, tubes with a non-return valve, and the nozzle holders (horn) with one or more nozzles. As well as fixed nozzle holders on the bumpers, there are also telescopically extending nozzle holders. The telescope improves the cleaning effect because it can assume an optimum spray position. In addition, when not in use, the nozzle holder can be concealed, e.g., inside the bumper.

Cleaning effect

The cleaning effect is chiefly determined by the cleaning pulse of the water jet. The decisive factors here are the distance between the nozzles and the headlamp lens, the size, the contact angle, the contact velocity of the washer-fluid droplets, and the amount of washer fluid. It is important that the nozzles be positioned correctly so that water jets properly cover the headlamps at all driving speeds.

References

[1] Commission Regulation (EU) No.1008/2010: Windscreen Wiper and Washer Systems.

Passenger-compartment climate control

Climate-control requirements

The system for controlling the climate in the passenger compartment of automobiles and commercial vehicles must fulfill many functions. Firstly, it must create a comfortable climate for all the occupants. To this end, cooled or heated air (depending on the outside temperature) is fed into the passenger compartment so as to meet the physiological needs of the occupants. A further function of climate control is to keep the passenger compartment free of unpleasant odors and pollutants.

Climate control also has the safety-related function of demisting and de-icing the windshield and side windows. Statutory requirements that vary from country to country must be met.

The comfort requirements and thus the functions of the A/C unit normally increase with the vehicle category. Regionally different climate-control requirements must also be taken into consideration. In Europe, for instance, vehicles with a system of climate control which is as unobtrusive as possible are preferred, whereas in the USA a cool, noticeable air current is desired.

Design and operating principle of the A/C unit

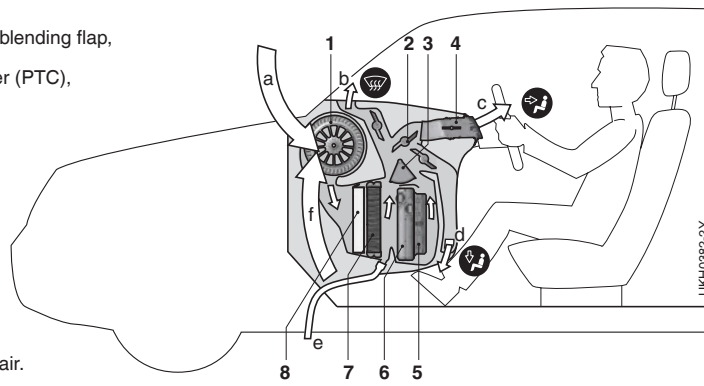
Air intake

The A/C unit is installed out of sight of the occupants under the vehicle's instrument panel. Fresh air is usually drawn in at the base of the windshield (Figure 1). Rainwater and snow are removed en route to the A/C unit. The outside air is drawn in by a radial fan and delivered through an opening in the bulkhead (partition between engine and passenger compartments). There is a further possibility of drawing in passenger-compartment air (recirculated air) through a second air inlet. The proportions of fresh air and recirculated air from the drawn-in air are controlled by flaps.

Fresh air is needed to supply oxygen to the passenger compartment. In recirculated-air mode, unpleasant odors and pollutants are prevented from entering the passenger compartment. In addition, power consumption is significantly reduced in recirculated-air mode, since the air to be temperature-regulated does not have to be heated and cooled so much in heating and cooling modes respectively. However,

Figure 1: Passenger-compartment climate control

1 Fan,
2 Vent flaps,
3 Temperature blending flap,
4 Air vent,
5 Electric heater (PTC),
6 Radiator,
7 Evaporator,
8 Filter.
a Fresh air,
b Air vent, windows (defrost),
c Air vent, ventilation,
d Air vent, footwell,
e Condensate drain,
f Recirculated air.



cold and thus dry fresh air must sometimes be used to prevent the windows from misting over due to condensation of damp passenger-compartment air.

Air cleaning
To clean the air, a sheet of filter fleece is pleated, i. e. folded, and preferably installed in the shape of a rectangular filter element before and after the fan in the A/C unit. The filter element is mainly positioned in such a way as to filter not only the fresh air, but also the recirculated air. The filter element has a limited retention capacity and must therefore be replaced. An activated-carbon layer can also be applied to the filter fleece (hybrid filter) in order to filter out particulates and pollutants.

Air cooling
The cleaned air is then cooled on the surface of an evaporator, which is now usually made entirely of aluminum. The air dissipates its energy proportionally to the evaporating refrigerant which is routed through the evaporator.

Because cold air can absorb less moisture than warm air, water is separated out as it cools and routed downwards along the evaporator. The water collects in the housing section under the evaporator, the condensate drip pan, and is discharged via the condensate outlet from the A/C unit to the atmosphere.

Air temperature regulation and distribution
After the evaporator the desired air temperature is set by a blending-flap system. To this end, the air is directed entirely or in part through a radiator (through which engine coolant flows) or – in pure cooling mode – past the radiator directly to the air vents. The air is continuously distributed to the two channels in any proportion by one or more temperature blending flaps. The cold and warm air currents are blended to suit demand in the downstream blending chamber. The air currents are blended in such a way as to cater for the comfort sensation of "cool head and warm feet". For this purpose air is blown out at different temperatures at the three main "Footwell", "Ventilation" (air vents in the instrument panel) and "Windows" vent levels, generating temperature stratification of up to 15 K.

The volume of air blowing out of the individual vents is continuously adjustable to any proportion by hand or by motor-driven vent flaps.

In further expansion stages of the A/C unit additional vents can be provided for the rear passengers or temperature regulation of the passenger compartment can be set differently for different seats. The first expansion stage is two-zone temperature regulation. The interior temperatures on the driver and front-passenger sides can be set independently of one another. Three-zone climate control caters for the rear-seat area as an additional climate zone, while four-zone climate control provides for individual climate control for driver, front passenger and two passengers in the back.

A new feature is individual adaptation of the footwell temperature, or "variable stratification". In this case, the outlet temperature in the footwell can be increased while the basic setting remains unaltered (for example for people with feet which are sensitive to cold).

Booster heater
In modern optimized-consumption internal-combustion engines the waste heat available in the engine coolant is reduced, and thus with it the heat supply in the heating circuit. In this case, an electric PTC booster heater is installed in the air passage after the radiator. By using PTC ceramic heating elements (PTC, Positive Temperature Coefficient), the booster heater automatically reduces its heat output at a defined limit temperature, since the resistance of the PTC elements increases abruptly. The PTC booster heater is therefore intrinsically safe.

The utilization of further heat sources, such as heat recovery from exhaust gas for example, is currently being developed.

Air routing
The temperature-regulated air is directed through plastic blow-channels to the relevant vents. Window and footwell outlets are not adjustable. The blow direction (nozzles) and the air volume can be individually adjusted at the air vents.

Supplementary A/C units
In luxury-class vehicles supplementary A/C units are sometimes use for the rear passengers. These are designed either, like the main A/C units, as heater-coolers or as pure coolers, and are usually supplied with recirculated air. Supplementary A/C units are installed, for example, in the center console, above the rear wheel arch, in the spare-wheel recess, or behind the rear seat bench.

Alternative A/C-unit concepts
The A/C unit described above sets the desired temperature by blending cold and warm air and represents the standard design used today. Alternative A/C-unit concepts may be used in which the air after the evaporator is routed directly through the radiator. A valve can be used to steplessly adjust the water flow rate through the radiator and thus the desired level to which the air flowing through is heated. The blending-flap system for blending cold and warm air is omitted.

Because of the thermal inertia of the water-filled aluminum radiator, desired changes to the air outlet temperature can only be completed slowly. To ensure that the passenger compartment is quickly cooled with the maximum air mass, these concepts normally utilize an additional air-side bypass past the radiator.

The advantage of this concept is the slightly more compact design of the A/C unit and the possibility of shutting off the radiator on the water side and thus avoid the unwanted situation of the air being heated in cooling mode by hot components in the A/C unit. The disadvantage is the expense of integrating a water valve.

Regulation of climate control
The climate of the passenger compartment is controlled by making inputs at the A/C operating unit, by adjusting the air vents and – in more luxurious vehicles – by means of various sensors. The variations range from direct mechanical activation of the A/C unit via Bowden cables and flexible shafts to fully automatic, processor-controlled air conditioning, where only the desired temperature is set.

The variables that can be influenced are blow-out temperature, air flow rate, air moisture, and air distribution to the different vents. The secondary variables are interior temperature including temperature stratification, air flow, acoustics at each seat, and misting status of the windows.

Most people find the following configurations pleasant: in summer mode cold air from the vents in the instrument panel, in winter mode warm air from the footwell and defrost vents (for keeping the windows clear), and in transition mode, heated air in variable proportions from the defrost, footwell and ventilation vents in the passenger compartment. Additional influencing factors can play a part in automatic air conditioning. The status and intensity of the Sun can be determined by solar sensors and included in the control operation. Air-quality sensors detect pollutants and odors in the ambient air and prevent them from entering the passenger compartment by closing the fresh-air inlets. Moisture sensors for the windshield enable action to be taken to prevent imminent window misting. The desired and actual passenger-compartment temperatures are compared by of temperature sensors.

Climate-control systems

Heating circuit
The waste heat from the engine is a source of energy which can be utilized without great expenditure to heat the passenger compartment. The hot coolant flows through the radiator in the A/C unit. The air passing through the radiator is heated.

In order to maintain the interior temperature in winter after a journey when the engine is switched off, it is possible to use an electric water pump, which forwards the coolant and makes the heat stored in the cooling system usable.

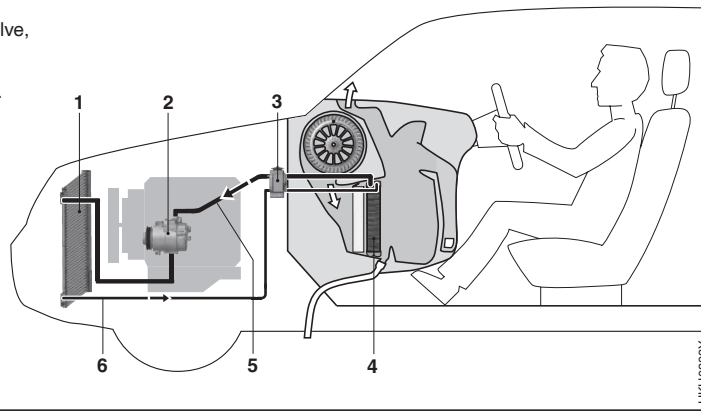
Refrigerant circuit
The function of the refrigerant circuit is to absorb the thermal energy of the air to be cooled in the evaporator and dissipate it at another point, outside the passenger compartment, to atmosphere. This is effected by using the cold-loss process in a closed refrigerant circuit (Figure 2). The main components of this circuit are the evaporator with expansion valve in the A/C unit, the condenser at the front of the vehicle, usually mounted directly in front of the coolant radiator, and the compressor secured to the engine block and driven by the engine.

The components are connected by metallic pipes which can contain flexible sections for decoupling. The compressor draws the gaseous refrigerant from the evaporator and compresses it, i.e. the pressure and temperature levels are greatly increased. In the condenser the hot refrigerant dissipates heat to the ambient air flowing through the condenser. This cooling condenses the vapor-state refrigerant. If the air stream is not enough to cool the refrigerant sufficiently, the radiator fan increases the air flow rate through the condenser. The refrigerant – now in liquid form and under high pressure – is atomized in the expansion valve and directed to the evaporator. The pressure level drops suddenly, as a result of which the refrigerant is highly cooled as it evaporates, and absorbs the heat of the air to be cooled in the A/C unit.

The evaporated refrigerant exits the evaporator in gaseous form and passes the expansion valve again, which, on the basis of the pressure and the temperature of the exiting refrigerant, alters the throttling cross-section on the inlet side so as to inject into the evaporator just that amount of refrigerant that can evaporate in the current operating state.

The cold gaseous refrigerant is drawn in again by the compressor and the circuit begins afresh.

Figure 2: Climate-control refrigerant circuit
1 Condenser with integrated refrigerant collector and dryer,
2 Compressor,
3 Expansion valve,
4 Evaporator,
5 Intake line,
6 Pressure line.



In order to prevent ice from forming on the air side of the evaporator, which would reduce the evaporator air cross-section, the minimum air temperature must not drop below 0 °C. To measure the air temperature after the evaporator, either an air-temperature sensor is used which regulates the compressor accordingly if the temperature drops below a fixed level, or the minimum intake pressure and thus the evaporator temperature are limited in the compressor. The refrigerant output can be adjusted between zero kilowatts (refrigerant does not flow through the evaporator) and a maximum output of approx. 8 kilowatts (for maximum refrigerant quantity and air flow rate).

Refrigerant
The fluorocarbon 1,1,1,2-Tetrafluorethane, which goes by the trade name R134a, is used as the refrigerant for vehicles with homologation before 2011-01-01 or vehicles where an existing type approval has been expanded (see Refrigerants, Operating fluids).

Because of this refrigerant's GWP (Global Warming Potential) of approx. 1,400 and the statutory provision which will come into force in Europe in 2011, which prescribes a refrigerant with a GWP of less than 150 for new vehicle models, an alternative refrigerant must be used here. There is currently still a lively debate regarding applicability of the alternative refrigerants. 2,3,3,3-Tetrafluoropropene with the trade name R-1234yf and carbon dioxide with the trade name R-744 are being pursued as alternatives. Starting at 2017-01-01 in Europe, all newly approved passenger cars must use a refrigerant with a GWP of less than 150.

Climate control for hybrid and electric vehicles

In vehicles with an engine start/stop function, it is necessary to keep the interior climate constant when the engine is switched off. For heating, an electric water pump can be used to utilize the residual heat from the engine circuit. For cooling, evaporators have been developed with a cold accumulator on the basis of a material which can store cold by means of a phase change and dissipate it again when stopped at traffic lights with the engine switched off. The interior climate can thus be kept at a pleasant level for up to two minutes.

In vehicles which run partly or entirely on electricity, a compressor is used which is electrically powered. Two solutions are available for heating. Either the water required for the radiator is electrically heated and supplied to the circuit by an electric pump, or the radiator is replaced by an electric heating element, the high-voltage PTC. Electric heater and electric compressor also open up the possibility of independent climate control. When the battery is sufficiently charged or connected to a charging station, the vehicle can be both preheated and precooled.

Because the power required for climate control can amount to several kilowatts and thus the range of an electric vehicle is reduced significantly, current developments are focused on increasing the efficiency of climate control.

Cabin filters

Function

Heating and air-conditioning systems in motor vehicles draw in air from the outside. After conditioning, the air passes into the passenger compartment together with any particles or gaseous pollutants it may contain. These air pollutants (Figure 3) can cause allergic reactions. Therefore, it makes sense to filter out particles and gases. A filter also reduces the amount of dirt that builds up on the fan, in the heating system, on the instrument panel, and on the inside of the windshield.

Depending on the requirement, either particle filters, charcoal filters, or biofunctional filters are used. Charcoal filters have the advantage that, besides filtering out particles, they can also prevent unpleasant odors from entering the passenger compartment. This is achieved by carefully embedding activated charcoal granules (up to 300 g/cm^2) on the particle-filter medium to eliminate aromatic substances. The activated charcoal also adsorbs ozone, benzene, and toluene, for instance. Biofunctional filters are furthermore used to prevent the growth of bacteria and to retain allergenic proteins.

Filter media

Whereas in the past filter media were primarily paper-based, the situation has now changed as a result of the greater demands placed on the filter system (removal of particles smaller than 0.001 mm). Present-day filter media generally consist of polyester or polypropylene-based nonwoven materials. The particle filter consists of three graduated fiber layers: the preliminary filter, the microfiber nonwoven layer with electrostatically spun microfibers, and the base nonwoven material. An activated charcoal filter also has an activated charcoal layer in addition to the three fiber layers. In a biofunctional filter a special coating on the filter medium provides the biofunctional effect.

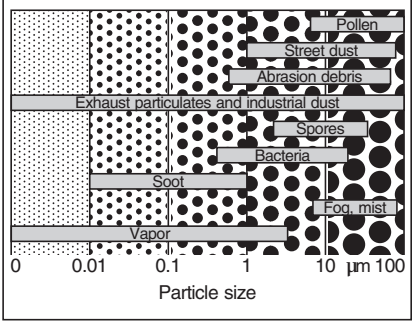
When the filter material is made into a filter, a number of specific parameters must be considered. There is a complex interrelationship between the filter material used, the depth of the filter-element folds, and the distance between the folds. The interaction between these parameters is important in determining the filter performance in service.

Design

Cabin filters must be optimized in three areas: filtration efficiency, pressure loss, and dust-storage capacity. The specifics of the application determine which of these three considerations is given precedence. The best filter may, for example, be a filter that achieves the highest possible filtration efficiency at the expense of a tradeoff in the filter service life. Another filter design may be one where a long service life is achieved by a high dust-storage capacity at a defined level of filtration.

The service life of filters used in passenger cars is around 20,000 km, i.e. with today's servicing intervals, they are replaced each time the car is taken for a scheduled service.

Figure 3: Prevailing particle sizes in road traffic



Auxiliary heater systems

Function and designs
Auxiliary heater systems use the fuel carried in the vehicle's tank to generate heat. Air heaters (Figure 4) are used to heat the passenger compartment directly. They transfer the heat generated during combustion to the cabin air. Coolant heaters (Figures 5, 6, 7) introduce heat into the cooling water and are thus suitable both for preheating engines and for heating the passenger compartment via the vehicle's heat exchanger with blower.

Air heaters are particularly efficient due to the direct transfer of heat to the air and are suitable for long periods of heating. Coolant heaters use the cooling circuit and the vehicle's existing air ducts and are therefore particularly suitable for covering the heat deficit of low-consumption drives and for preheating the passenger compartments and engines of passenger cars.

Further advantages of auxiliary heater systems are:
– Ice-free windows thanks to preheating,
– Optimized heating comfort,
– Use of a truck sleeping cabin even in winter,
– Less wear and lower emissions when starting with a preheated engine (with coolant heater only),
– Operating temperature of the catalytic converter reached more quickly (with coolant heater only).

Of increasing importance is their use in passenger cars with deficient heat output from waste heat caused by increased drivetrain efficiency by means of electrification. Particularly in plug-in hybrid vehicles – but also in purely battery-powered vehicles – fuel-powered heaters, on account of their great efficiency in converting chemically bound energy into heat, ensures that the range in electric driving mode is not limited in the case of the storage capacity of the traction batteries being specified.

Air heaters
Application
Air heaters are primarily used in trucks and commercial vehicles. Their main advantages are low costs, faster installation, and lower power and fuel consumption. A dramatic increase in their use was witnessed in the USA with the introduction of legislation prohibiting the running of the engine to provide heat during breaks in driving since heaters can heat the cabs with clearly lower fuel consumption.

Operating principle
Air heaters (Figure 4) operate independently of the vehicle's own heat balance. A combustion-air blower draws the air required for combustion from the surrounding atmosphere, and blows it into the burner. An electric metering pump delivers diesel or gasoline as fuel via an evaporator (component not visible in Figure 4) to the burner, where it mixes with combustion air, is ignited by a sheathed-element glow plug, and burns. The heating-air blower draws in the fresh air to be heated, blows it through a heat exchanger, and into the vehicle cabin.

In cold countries a heat output of approx. 4 kW is required. In temperate climates units with approx. 2 kW are sufficient.

An important safety aspect is that combustion air and exhaust gas from the cabin air must be kept entirely separate. This prevents exhaust gas from entering the cabin.

The heat output is regulated by varying the quantity of fuel, combustion air and fresh air. If the temperature measured by the room-temperature sensor deviates from what is set at the operating element (e.g. radio remote control), the output of the heater is adapted until the desired temperature is reached once again. An overheating sensor registers unacceptable temperatures through defects or malfunctions that occur (e.g. overheating due to blocked heating vents) and switches off in good time if necessary. A flame sensor determines by means of the heat-exchanger temperature whether the burner has ignited and detects when the flame goes out.

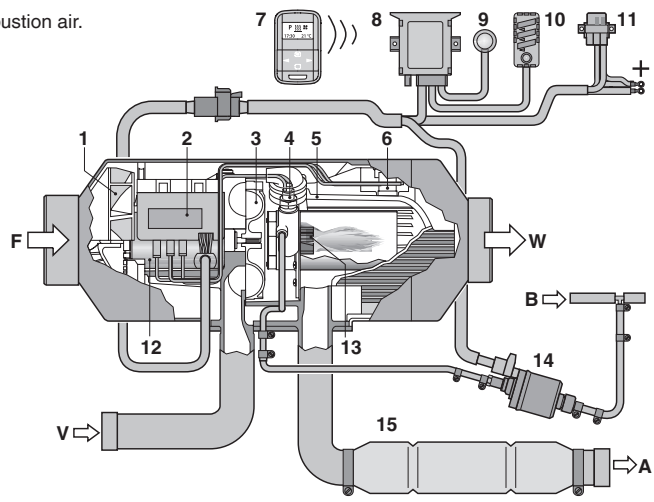
Installation
Air heaters for commercial vehicles can usually be fitted directly inside the cab. In trucks, for instance, the preferred locations are in the co-driver's footwell, on the rear cab wall, under the bunk bed, on the outside of the cab wall, or in a stowage compartment. The exhaust pipe always runs under the floor (either into the wheel arch, or out of the rear cab wall).

Most sensor units in the fuel tanks of passenger cars and trucks already have a connection available for supplying fuel to the air heater. If necessary, an extra tank sensor unit is fitted. A fuel reserve for driving is guaranteed by the immersion depth of the fuel-tank extraction fittings or the vehicle electronics, which switches off the heater from a minimum fill level.

Coolant heaters
Application
Coolant heaters up to approx. 5 kW heating capacity are used both in passenger cars and in trucks. They use the cooling circuit and already existing heat exchangers to heat the engine and cabin, as well as the available fans, air ducts, flaps, and vents. They are thus easy to retrofit. Heaters up to approx. 12 kW are used in trucks to preheat the engine. The large body surface area and high interior volume of city and tour buses mean that they require heating capacities of up to 35 kW.

The main types are auxiliary and booster heaters for passenger cars. These heaters are integrated directly in the cooling-circuit feed line between the engine and the heat exchanger of the passenger

Figure 4: Air heater
1 Heating-air blower, 2 ECU, 3 Combustion-air blower, 4 Sheathed-element glow plug,
5 Heat exchanger, 6 Combination sensor (overheating and flame sensor),
7 Radio remote control (mobile part), 8 Radio remote control (stationary part),
9 Button, 10 Room-temperature sensor, 11 Fuse holder, 12 Electric motor,
13 Burner, 14 Metering pump, 15 Exhaust muffler.
F Fresh air from vehicle cabin,
W Hot air to vehicle cabin,
A Exhaust gas,
B Fuel,
V Combustion air.



UKH0378-4Y

cell (Figure 5). They utilize the existing devices: vehicle heat exchanger with blower, air flaps and vents.

Quick-starting heaters are used to meet high demands with regard to heating-up speed; these reach their full heat output level after approx. 30 s. The required acceleration of fuel vaporization is achieved by heating the vaporizing apparatus electrically during the starting sequence.

Auxiliary heaters for preheating engine and passenger compartment before starting driving

Heat is generated through combustion of fuel along similar lines to air heaters. The generated heat is transferred to the cooling water flowing between the water jacket and heat exchanger (Figure 7) of the heater.

The simplest method of integrating them in the vehicle's cooling system is to

Figure 5: Coolant heater, integrated in cooling circuit
(Standard type)
1 Internal-combustion engine, 2 Water pump,
3 Coolant heater, 4 Vehicle heat exchanger with blower.

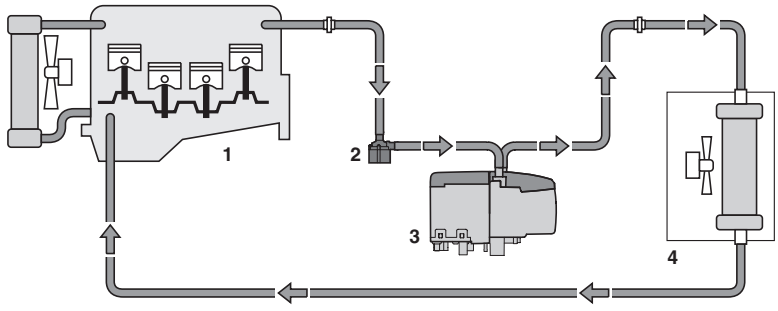
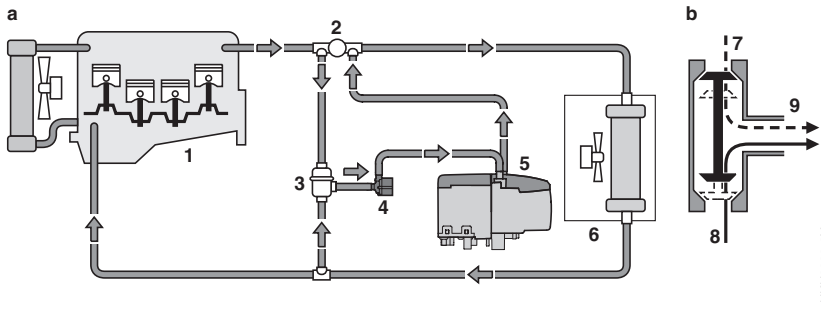


Figure 6: Split cooling circuit for priority heating of passenger compartment
(Split into secondary and primary cooling circuits)
a) Layout, b) Thermostat connections.
1 Internal-combustion engine, 2 Non-return valve, 3 Thermostat, 4 Water pump,
5 Coolant heater, 6 Vehicle heat exchanger with blower,
7 Return flow from vehicle engine, 8 Return flow from heat exchanger, 9 Connection to heater.
Separation by thermostat and non-return valve into small water circuit (priority for interior heating) and large water circuit (with motor).

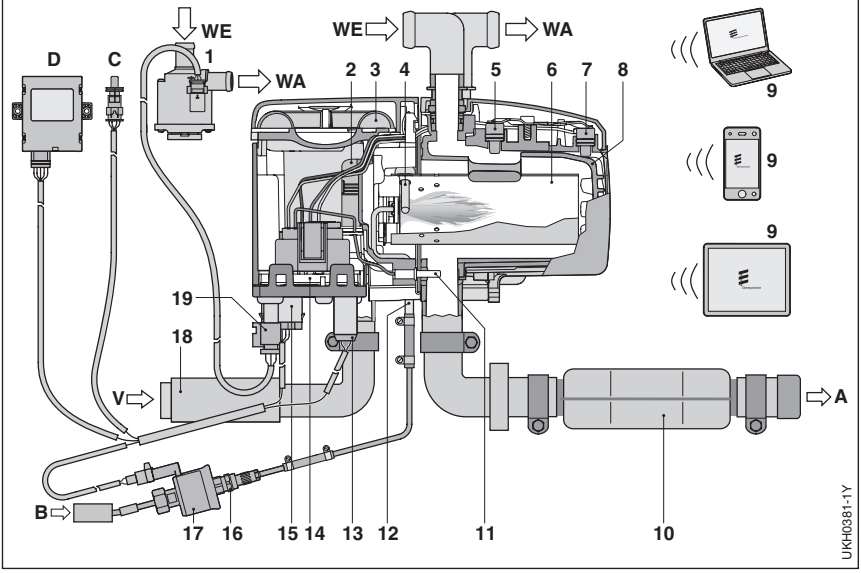


connect them in series between the engine and the heat exchanger in the primary circuit (Figure 5). A disadvantage with engines over approx. 2.5 l is that engine heating makes heating the passenger compartment much slower. In this case, the passenger cell can be heated as a matter of priority via an additional small circuit (Figure 6). This offers, irrespective of the engine, the advantage of being able to use the heating also for short-distance service without discharging too much the starter battery, which during the starting phase powers the glow element, the burner fan, the water pump, and the metering pump. Flow passes through the engine only after the thermostat opens when a high water temperature is reached. The non-return valve prevents the circuit from being bypassed via the thermostat.

The water pump (Figure 7) in auxiliary heaters pumps cooling water to the heating-system heat exchanger, where it is heated by heat transfer by the burner exhaust gas. On the one hand, the heated water preheats the vehicle engine; on the other hand, the vehicle's own heat exchanger (Figures 5 and 6) dissipates the heat. The heated air is blown controllably through the existing ventilation system into the vehicle cabin.

Figure 7: Coolant heater

1 Water pump, 2 Combustion-air motor, 3 Fan wheel, 4 Glow element, 5 Temperature sensor,
6 Burner, 7 Temperature sensor, 8 Heat exchanger,
9 Web-based operating elements (laptop, smart phone, tablet),
10 Exhaust muffler, 11 Flame sensor, 12 Fuel connection,
13 Plug-in connection, blower-flap control module, 14 ECU, 15 Connection to heater,
16 Pot strainer, fitted in metering pump, 17 Metering pump, 18 Combustion-air muffler,
19 Water-pump connection,
A Exhaust gas,
B Fuel,
C Diagnosis connection,
D Blower-flap control module,
V Combustion air,
WA Water outlet,
WE Water inlet.



Control for manual immediate operation and for preprogramming switch-on times and heating duration is performed via radio remote control, via the Internet for example by means of a smart-phone app, or via a menu in the vehicle's own A/C operator unit.

Booster heaters for covering heat deficits while driving
Efficiency-optimized drivetrains, such as highly turbocharged diesel and gasoline engines with small displacements and hybrid drives, generate on account of their efficiency insufficient waste heat to heat the passenger compartment. Booster heaters are used to make up this heat deficit. They operate only at outside temperatures below +5 °C and when the engine is running. They therefore do not require their own water pump.

Booster heaters can be upgraded to an auxiliary heater by retrofitting a water pump and a control element with ECU.

When used in an electric vehicle or in a plug-in hybrid vehicle, the fuel-powered heater can cover the entire heat demand.

Installation
To avoid heat losses through long cooling-water hoses, coolant heaters are usually installed in the engine compartment. The confined installation conditions necessitate a very compact construction. Installation is simplified by integrating all the components required to generate and deliver heat in the heater unit (Figure 7).

Regulations
All auxiliary heaters and booster heaters run on fuel have type approval in accordance with ECE Regulations No. 10 [1] and No. 122 [2].

If a heater with this type approval is retrofitted in line with the manufacturer's installation instructions, no inspection/approval by an expert or inspection body is prescribed. The installation of auxiliary heaters in vehicles used for international transport of hazardous goods is governed by the European ADR agreement (Accord européen relatif au transport international des marchandises Dangereuses par Route). The heater must be switched off for safety reasons before the vehicle enters a hazardous area (e.g. refinery or filling station). The heating also switches off automatically as soon as the vehicle engine is shut down or an ancillary (e.g. auxiliary drive for discharge pump) is switched on.

Regulations in the USA
The heater requires CARB approval (emissions certificate) for installation and operation in medium and heavy commercial vehicles.

References
[1] ECE R 10: Regulation No. 10 of the Economic Commission for Europe of the United Nations (UN/ECE) – Uniform provisions concerning the approval of vehicles with regard to electromagnetic compatibility.
[2] ECE R 122: Regulation No. 122 of the Economic Commission for Europe of the United Nations (UN/ECE) – Uniform technical provisions concerning the approval of vehicles of categories M, N and O with regard to their heating systems.

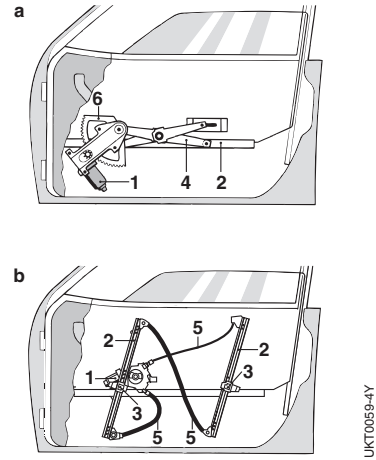
Comfort and convenience systems in the door and roof areas

Power-window systems

Power windows have mechanisms that are driven by electric motors. There are two types of design in use (Figure 1):

– Arm window lift: The window-lift-drive pinion engages with a quadrant gear, which is connected to a rod linkage. This system is highly efficient. On the down side, the system requires a large amount of space for fitting and is heavy.
– Flexible cable window lift: The window-lift drive turns a cable reel, which operates a flexible cable mechanism. Systems with two guide rails are used for the most part in front doors. Systems with one guide rail are preferred in rear doors. A crucial advantage of these designs is their excellent ability to guide the window glass.

Figure 1: Power-window systems
a) Arm window lift, b) Flexible cable window lift.
1 Power-window drive,
2 Guide rail,
3 Driver,
4 Rod linkage,
5 Drive cable (Bowden cable),
6 Quadrant gear.

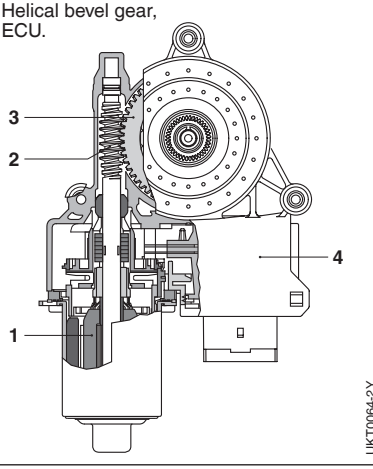


Power-window drive

The power-window drive consists of a DC motor with downstream reduction gearing (Figure 2). For comfort and safety reasons, many power-window drives are equipped with an ECU. The gearing is designed as a helical bevel gear to create the necessary self-inhibiting effect which prevents the window from opening inadvertently of its own accord or being opened by force. Special tribological measures and the magnetic-circuit configuration of the DC motor further improve the drive's self-inhibiting properties. Space limitations inside the door make narrow (flat) drive units imperative.

Dampers integrated in the gearing mechanism provide good damping characteristics at the window-glass end positions.

Figure 2: Power-window drive with integrated ECU
1 Power-window drive motor,
2 Worm,
3 Helical bevel gear,
4 ECU.



Electronic control unit (ECU)

Two different operating modes are used. In manual mode the power-window drive is operated using a switch which is pressed over the full extent of window travel. In automatic mode the window glass is opened or closed by a short pressing of the switch.

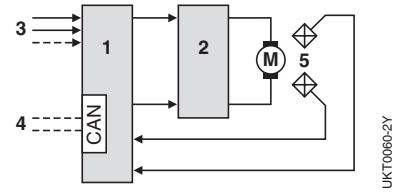
A force-limitation device is required by law to avoid injuries in the course of au-

tomating closing. In Germany, paragraph 30 of the StVZO [1] (Road Traffic Licensing Regulations) stipulates that the force-limitation device must be effective when the window is moving upwards within a range of 200 to 4 mm from the upper edge of the window aperture. The power-window drive includes integral Hall sensors to monitor motor speed during operation (Figure 3). If a drop in speed is detected, the DC motor's direction of rotation is immediately reversed. The window closing force, measured for a spring rate of 10 N/mm, must not exceed 100 N. To be able to close the window safely, the force-limitation device is automatically deactivated before the window glass enters the door seal. The window position is monitored over its entire range of movement.

Depending on the vehicle topology, the electronic control may be accommodated in a central ECU or decentrally in the door, preferably directly in the power-window drive. Decentralized electronics can be networked via a LIN-bus interface. The advantages of such a solution are fault diagnosis of the electronics and reduction of the amount of wiring.

Figure 3: Power-window control unit featuring electronic force-limitation device

1 Microcomputer,
2 Relay output stage,
3 Control commands,
4 Networking via CAN,
5 Hall sensors.



Sunroof systems

Design variations
Power-sunroof drives can combine tilting- and sliding-sunroof functions. Three different types of sunroof systems may be used:
– The simple sliding/tilting sunroof: A panel (frequently made of glass) can be tilted or slid to the rear underneath the roof skin.
– The large sunroof or panoramic sunroof: This system utilizes up to three sliding-sunroof drives with different adjustment options (Figure 1, e.g. adjusting a glass panel in the front and rear areas, or with a fixed glass panel and an electrically adjustable roller sun blind).
– The spoiler sunroof: Here the glass panel is tilted and then moved to the rear over the roof skin (hence the term “spoiler”).

Power-sunroof drives
Sunroofs and roller sun blinds are primarily adjusted by means of torsion- and pressure-resistant control cables or plastic toothed straps with mechatronic drives. These are located in the roof between the windshield and the sliding sunroof or at the rear of the vehicle.

In the event of an electric-adjustment failure, the sunroofs can be closed by means of a hand crank provided on the drive.

The drive consists of a permanent-field DC motor with helical gear drive, a power output of approximately 30 W and an ECU. The motor is protected against thermal overload by software thermal protection.

Electronic control
The electronic control system operates with a microcomputer responsible for evaluating incoming signals and monitoring sliding-sunroof position.

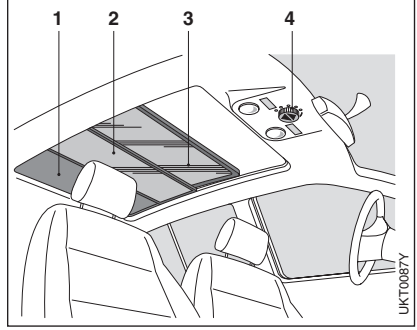
Control is provided by means of exact position pulses. The pulses come from the ring magnet (up to 12-pole), which is mounted on the motor armature. They are sensed and evaluated by two Hall-effect switches. As well as exact positioning, it is also possible to effect closing-force limitation in order to avoid injuries during automatic operation when both opening and closing.

The drive is actuated via external signal inputs using pushbuttons, switches, and analog or digital preselector switches (e.g. potentiometers, digital sensing via Gray code).

Control can be integrated in the vehicle’s bus system (CAN, LIN) and diagnostic information can be output. Electronic control offers the opportunity to easily implement a wide range of functionalities and comfort and convenience functions (e.g. preselectable position control, closing via remote control, automatic closing in the event of rain, combination with the central locking system).

References
[1] German Motor Vehicle Safety Standards (StVZO) – § 30 Beschaffenheit der Fahrzeuge (State of vehicles).

Figure 1: Panoramic sunroof with operating facility
1 Sliding roller sun blind,
2 Fixed or sliding glass panel in rear area,
3 Sliding glass panel in front area,
4 Operating unit.



Comfort and convenience functions in the passenger compartment

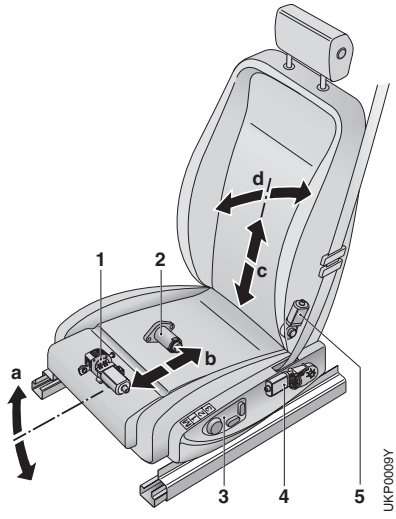
Electrical seat adjustment

On the one hand, the electrical seat adjustment ensures greater operating convenience than mechanical operation, and, on the other hand, it is also preferred on account of its lower space requirement or in the case of poor accessibility of the various adjustment planes.

The four main adjustment planes feature longitudinal adjustment of the full seat, tilt adjustment of the seat backrest, and height and tilt adjustment of the seat cushion.

Figure 1: Electromechanical seat adjustment

1 Electric motor for longitudinal adjustment of the seat (gearing is part of the seat structure),
2 Electric drive for height adjustment,
3 Operating unit (on the seat or alternatively in the door),
4 Electric drive for tilt adjustment of the seat cushion,
5 Electric drive for tilt adjustment of the seat backrest,
a Tilt adjustment of the seat cushion,
b Longitudinal adjustment of the full seat,
c Height adjustment of the seat,
d Tilt adjustment of the seat backrest.



Drives

Electric motors with different gear variants are used (e.g. lifting and rotating spindles, planetary gears, and two-stage spur gears) are used as drives for electrical adjustment.

To simplify entry and exit, the drives can be designed in such a way that adjustment is completed particularly quickly (easy entry). This function is particularly useful for gaining access to the second row in a two-door vehicle.

Position detection

Drives with position detection are increasingly being used to enable the seat position to be stored individually for different drivers and to retrieve and adjust this position (memory function). Position detection is effected either via Hall generators or via sensorless evaluation of the current signals from the electric motor (also called Sensor Less Control, SLC).

The memory function simplifies and speeds up adjustment of the seat position for an individual driver after a change of driver.

Further adjustment options

Aside from the four main adjustment planes for the driver and the front passenger, a wide range of additional adjustments can be performed electrically such as seat-cushion depth, head restraint, and lumbar support.

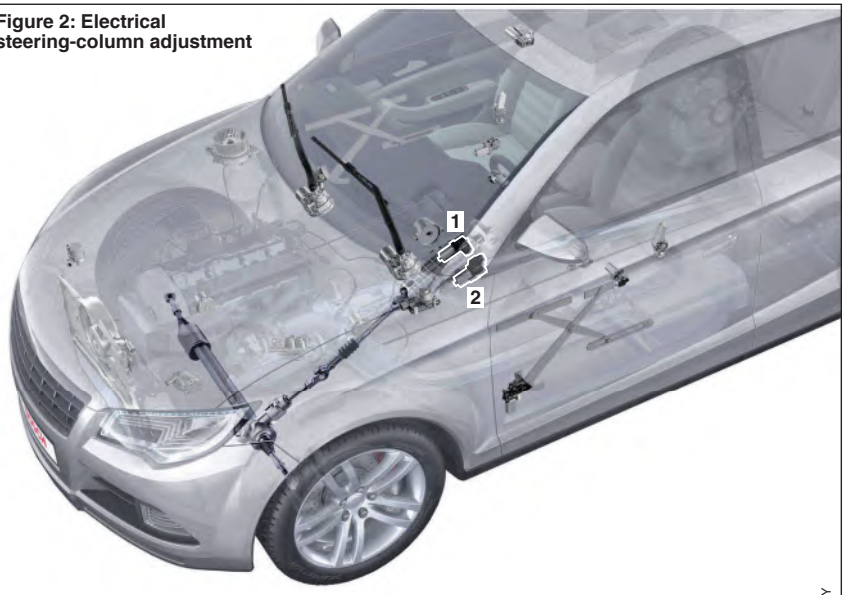
Electrical adjustments are increasingly also being performed for the second and third seat rows in larger vehicles (e.g. off-road vehicles). A significant motivation for this is to make it much easier and more convenient to load larger objects.

Electrical steering-wheel adjustment

A further comfort and convenience function which is often combined with electrical seat adjustment is electrical steering-wheel adjustment. Here both the longitudinal clearance and the height of the steering wheel can be infinitely adjusted via a drive similar to the spindle drives for seat adjustment.

This function is particularly useful in combination with the memory function. In this way, for example, entry and exit are made easier by extending the clearance between the driver's seat and the steering wheel prior to entry and exit. During driving the positions of the seat and steering-wheel adjustments individually preprogrammed for the driver are set.

Figure 2: Electrical steering-column adjustment



1 Electrical steering column height adjustment drive,
2 Electrical steering column longitudinal adjustment drive.

Safety systems in the motor vehicle

Phases of vehicle safety

Vehicle safety can be broken down into the categories of active safety and passive safety. Active safety systems help to prevent an accident. Passive safety systems are used when an actual accident occurs.

Vehicle safety is divided into five phases (Figure 1). Active safety includes normal driving, in which safety systems issue warnings, provide recommendations, or make corrections (phase 1 and phase 2). In the event of a collision passive vehicle safety intervenes in a protecting (phase 3 and phase 4) and in the post-crash phase in a life-saving capacity (phase 5).

Different safety systems are available, depending on the vehicle equipment. Owing to its effectiveness in accident avoidance, driving-dynamics control (Electronic Stability Program, Electronic Stability Control, ESC) for example has established itself in all vehicle classes and has been required by law in Europe since 2012 for newly registered vehicles.

Normal driving operation

Belt-reminder system

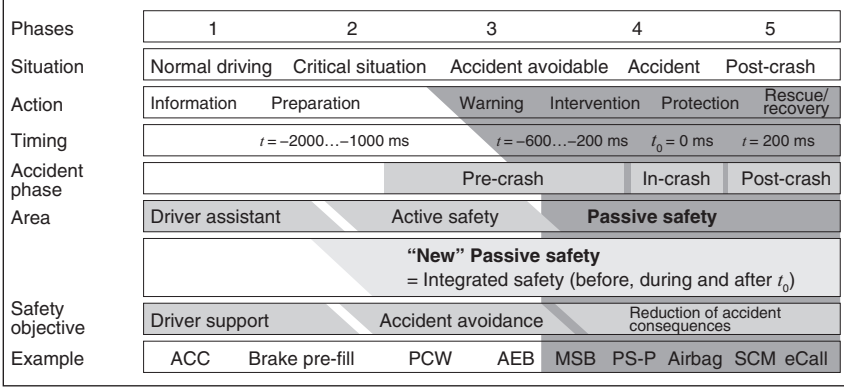
If the driver or front passenger does not have their seat belt fastened, there is a visual and acoustic warning signal (SBR, Seat Belt Reminder, belt-reminder system, Figure 2a). Sensing is performed for example by means of a seat-occupancy mat (see Occupant sensing). The presence of further occupants in the rear passenger compartment is optionally detected via sensors. In this case, the term used is Enhanced Seat Belt Reminder (ESR).

Child seats

The use of child seats is required by law. Provisions also stipulate that the front-passenger airbag must be deactivated if a rear-facing child seat is placed on the front passenger seat (COP, Child Occupant Protection). The airbag can be manually or automatically switched depending on regional variation. The corresponding function lamp lights up when the airbag is deactivated (Figure 2b).

Figure 1: Vehicle-safety phases

ACC: Adaptive Cruise Control
PCW: Predictive Collision Warning
MSB: Motorized Seat Belt (electro-mechanical belt pretensioner)
PS-P: Presafe-Pulse, triggering before t_0
Brake pre-fill: Application of brake pads
AEB: Automatic Emergency Braking
SCM: Secondary Collision Mitigation
eCall: emergency Call

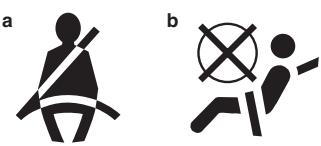


Critical situation
If a critical situation ensues for example in response to full braking or in the event of braking on a different surface, targeted brake interventions on individual wheels and interventions in the engine management bring about a stabilization of the vehicle (driving-dynamics control). Another example of how the vehicle is prepared for a potential accident is the pre-filling of the brake system (brake pre-fill). The braking effect starts earlier once the brake pedal is pressed, resulting in a shorter braking distance.

However, physical limits are placed on the effectiveness of these safety systems. An accident is unavoidable if these limits are exceeded.

Before the crash
For improvement of the deployment function of the occupant-protection systems and advance detection of the type of impact (pre-crash detection), microwave radar, ultrasound or lidar sensors are used to measure relative speed, distance, and angle of impact for frontal impacts. If distances are too short at a corresponding speed, optical and acoustic warnings can be issued or the brake can be activated (AEB, Automatic Emergency Brake, see Emergency brake systems). Electromechanical reversible seat-belt pretensioners are used with pre-crash sensing. The fact that they are reversible means that in a dangerous situation they can already be tightened prior to a possible collision. In this way, belt slack is already eliminated before the crash, enabling the occupants already at an early stage to participate in the vehicle's deceleration during a braking operation.

Figure 2: Warning lamps in the instrument cluster
a) Symbol lights up if seat belt is not fastened.
b) Front-passenger airbag deactivated.



During the crash
The crash starts at t_0, the moment of first vehicle contact (in-crash) with the obstacle, and ends with the vehicle at a standstill. In this time window all the relevant retention systems are activated (see Occupant-protection systems) and if necessary emergency braking is initiated; this is intended to avoid a potential secondary accident or to mitigate the accident severity.

After the crash
Event Data Recorder
Vehicles with airbag systems store data relevant to the accident, e.g., impact accelerations, belt-buckle states or airbag deployment times, in a non-volatile memory of the airbag ECU (EDR, Event Data Recorder, crash recorder). The volume of data varies depending on the vehicle manufacturer and records a period of around 100 ms before and after the impact.

This enables the accident to be analyzed to a certain extent. For example, accidents with airbag deployment can be distinguished from near-accidents (near misses) and also accidents without airbag deployment. For another thing, the data permit a partial reconstruction of the vehicle situation before the accident for accident-research purposes. In addition, statutory requirements are thereby fulfilled. Thus, in the USA the use of an event data recorder in new vehicles has been required by law since 2012.

eCall
eCall (abbreviated form for emergency call) is an automatic emergency-call system for motor vehicles prescribed by the European Union which since April 2018 has been required to be installed in all new models of passenger cars and light commercial vehicles.

The airbag ECU activates, aside from hazard warning lights, door openers, deactivation of the fuel pump, and isolation of the battery, an emergency call via a carphone installed in the vehicle. The emergency-call center automatically receives information on the exact location of the vehicle and can if necessary talk to the occupants and initiate recovery and rescue operations.

Active safety

The aim of active safety is already in advance to prevent an accident from happening. The level of technical vehicle equipment helps to recognize dangers quickly in normal driving conditions and to avoid accidents. Active safety therefore already begins with a good view of the road traffic even at dusk/dawn and at night. The range of vision was virtually doubled when Bilux lights were replaced by halogen lamps in the 1970s. The introduction of xenon and LED headlights were further milestones.

Well-conceived operating concepts e.g. for the infotainment system via the multifunction steering wheel and the depiction of information in a head-up display enable the driver to keep their eyes on the road and the traffic. Even this is a contribution to active safety.

In the field of driver assistance the brake booster and power steering for example offer the potential to avoid accidents.

But active road safety also includes the driver's driving qualifications (e.g., driving-safety training), their traffic literacy and competence, their traffic-appropriate behavior, and their own responsible actions and initiative.

But, generally speaking, active safety refers to systems which actively intervene in the driving dynamics and thus actively support the driver. These systems are also called driver-assistance systems.

Examples of active safety systems

Active safety systems require information about the vehicle's driving and operating state (e.g., wheel speeds, yaw rate, steering-wheel angle) and also information about the vehicle's environment or surroundings (e.g., distance to the vehicle traveling in front). The advances in sensor technology led to the development of numerous active safety systems.

1978 saw the introduction of the first antilock braking system (ABS) ready for series production, a system which prevents the wheels from locking during heavy braking and thus keeps the vehicle steerable (see Wheel-slip control systems). Further development led to driving-dynamics control (Electronic Stability Control, ESC), in which the antilock braking system and the traction control system are integrated and which additionally detects the vehicle's swerving/skidding tendency and counteracts this tendency with targeted wheel-individual brake interventions and interventions in the engine management.

Adaptive Cruise Control (ACC), initially developed as a comfort and convenience system, has developed further in the direction of an active safety system. It relieves the strain on the driver by automatically adapting the speed to the vehicle traveling in front and where necessary brakes the vehicle to a standstill. In this way, the safety distance is constantly maintained.

Emergency Brake Assist performs an emergency-braking operation if it detects an object in front of the vehicle.

Lane Keeping Support warns the driver for example by vibrating the steering wheel if the vehicle leaves the lane without a turn signal having been activated. If the driver fails to react, the system can automatically steer the vehicle back to the original lane.

Blind-spot assistants and lane-change assistants monitor the areas to the sides of and behind the vehicle and warn the driver when a vehicle is approaching. These systems thus warn when a passing (overtaking) maneuver cannot be safely carried out.

Importance of active safety

Safety in motor vehicles was previously confined to passive safety. Occupant-protection systems have in the meantime become a matter of course; advances in these systems are more likely to occur in the background. Active safety systems are constantly subject to further development and new systems are always arriving. They are deployed initially in premium vehicles and then also appear in the lower vehicle classes. New active safety systems are vigorously promoted by the vehicle manufacturers.

Passive safety

Passive safety in a car increases occupant protection in the actual event of an accident. There are 150 milliseconds which decide in the event of a car accident whether the outcome remains confined to damage to property or escalates to serious injuries to the occupants. If an accident occurs, numerous vehicle components can be crucial to survival. From the passenger safety cell extending across a stable body, crumple zone, energy-absorbing steering column and headrests through to airbags and seat belts with belt pretensioners (see Occupant-protection systems), passive safety is aimed at mitigating the consequences of an accident. Many of these components are required by law and are therefore standard equipment in the vehicle.

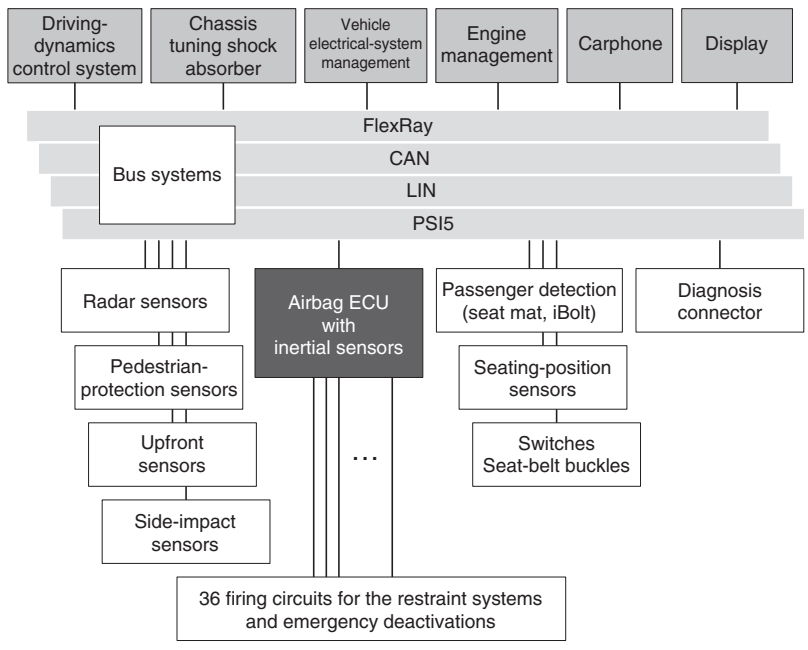
Whereas active safety aims at avoiding an impending accident entirely, passive safety is intended to mitigate the consequences of an unavoidable accident. By linking with driver assistance and active safety, new enhanced passive-safety functions provide for more comprehensive occupant protection (see Figure 1).

System architecture, passive safety
The passive-safety functions are specific to the manufacturer, vehicle, and equipment specification. Figure 3 shows a possible architecture.

The airbag ECU collects information on the speed, temperature, belt-buckle states, the seating position of the occupants, and registers activation of the ignition. Diagnoses, driving states, and accidents are recorded in the vehicle and stored in an event recorder. Data can be read via the diagnosis connector and if necessary software updates can be performed.

Inertial-sensor signals, i.e., accelerations and yaw rates in all three directions, are also made available for other systems. In this way, spring-damping systems or

Figure 3: System architecture, passive safety



even brake and steering systems can be activated with these signals to optimize and stabilize the vehicle in terms of driving dynamics.

The information already in the lead-up to a crash – e.g., on the relative speed, moment of impact of potential collision objects – comes from a radar, lidar or video sensor system. This enables seats or reversible belt pretensioners for example to be preconditioned for the crash situation or warning signals to be issued.

The vehicle electrical-system management uses the signals from the airbag ECU to control the airbag warning lamps, the status of the seat belts and seat occupancy, the hazard-warning lights, and emergency door opening.

An emergency call is sent via the carphone in the event of a crash; there is also the possibility of communicating information on medical treatment to the emergency services in advance. The engine management switches off the fuel pump after a crash; if necessary, the battery is isolated.

Legal regulations and consumer tests

Legal regulations

Requirements with regard to passive safety in vehicles are laid down in legislation. These cover safety of drivers, passengers, and in particular children and pedestrians.

Safety requirements laid down in the EU are for example the provisions pertaining to head-on- and side-impact protection for the vehicle occupants and pedestrian protection. These are reflected in UN/ECE regulations, which apply to the EU but have in the meantime also be adopted by many non-European countries. Other countries follow these regulations to a large extent (exception: USA).

The USA has the Federal Motor Vehicle Safety Standards (FMVSS), which constitute the technical regulations pertaining to vehicle safety and are legislative in character.

Consumer tests

The country-specific NCAP (New Car Assessment Program) is intended to provide car buyers and car manufacturers with a realistic and independent assessment of the safety features of some of the biggest-selling vehicles. Standardized crash and rollover tests of vehicles are conducted for this purpose. The vehicles are occupied by dummies during the crash tests. The safety assessment is derived from the measured values recorded on the dummies. To assess the safety of child seats, child dummies are placed in the child seats recommended by the vehicle manufacturer.

Thus, Euro NCAP is a voluntary European program for assessing vehicle safety which is supported by the European Commission and a number of European governments, and by automobile clubs and consumer organizations. Euro NCAP publishes safety reports on new cars and awards a star rating, based on the performance of the vehicles in various crash tests, including head-on, side and pole impact and collisions with pedestrians [1]. The highest rating is five stars. The requirements imposed on the star rating increase in parallel with vehicle development so that a 5-star car ten years ago is not comparable with a 5-star car today.

The test conditions are constantly updated, the test program extended to include additional tests, and the assessment criteria tightened. More recent findings concerning accidents and their causes are reflected in new test-condition requirements. Where originally only the results from the crash tests were used for assessment, assessments of active safety (e.g., lane departure warning, emergency brake assist) are increasingly being included in the overall assessment.

References

[1] www.euroncap.com/en/euro-ncap/how-to-read-the-stars/

Occupant-protection systems

In the event of an accident, occupant-protection systems are intended to keep the accelerations and forces that act on the passengers low and thereby lessen the consequences of the accident.

Seat belts with belt pretensioners provide a substantial proportion of the protective effect, absorbing the majority of the kinetic energy of the occupants. Together with a front airbag, the energy in the crash must be reduced in such a way that the occupant is not propelled up against the steering wheel or the dashboard.

In order to achieve optimum protection, the response of all components of the occupant-protection system must be adapted to one another. This is made possible by suitable sensors and high-speed signal processing. Optimum occupant protection against the effects of an impact is obtained through the precisely coordinate interaction of electronically fired, pyrotechnically inflated front airbags and seat-belt pretensioners. To maximize the effect of both protective devices, they are activated by the airbag ECU (triggering unit) installed in the passenger cell.

Figure 1 provides an overview of the occupant-protection systems in a passenger car.

Functions

Environment sensing

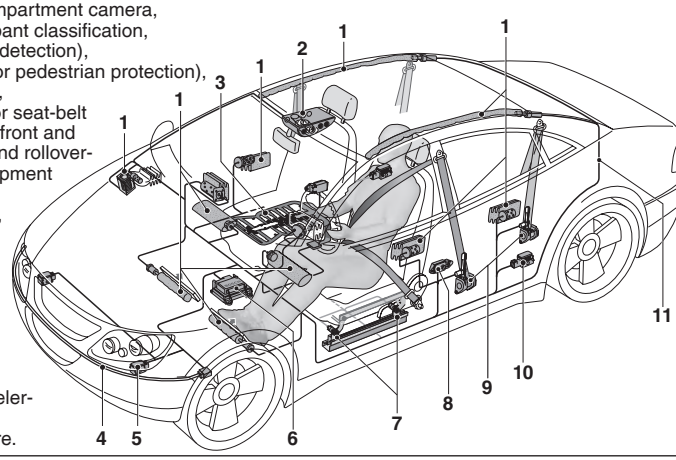
The ECU detects the different types of impact (e.g., head-on, oblique, offset, pole, rear-end, side impact) in the vehicle longitudinal direction (x-axis) and vehicle transverse direction (y-axis) with the aid of integrated acceleration sensors or structure-borne noise sensors. From their signals the crash severity and crash type to be expected is determined so that the appropriate retention systems can be activated at the right time.

Peripheral acceleration and pressure sensors installed on the vehicle body are connected via leads to the ECU. They facilitate earlier detection of vehicle contact and hence improved performance with regard to deployment of the retention systems. Especially in the case of side impacts with a small crumple zone, high-speed pressure sensors installed in the door have an advantage over acceleration sensors.

Rollover detection is performed by yaw and acceleration sensors. Aside from the yaw rate about the longitudinal axis, the vehicle lateral acceleration (y-axis) and the acceleration in the direction of the vertical axis (z-axis) are sensed.

Figure 1: Occupant-protection systems

1 Airbag with gas inflator,
2 Passenger-compartment camera,
3 OC mat (occupant classification, seat-occupant detection),
4 Tube sensor (for pedestrian protection),
5 Upfront sensor,
6 Central ECU for seat-belt pretensioners, front and side airbags, and rollover-protection equipment with integrated rollover sensor,
7 iBolt,
8 Peripheral pressure sensor,
9 Seat-belt pretensioner with propellant charge,
10 Peripheral acceleration sensor,
11 Bus architecture.



The analysis algorithms in the ECU read in the data from the sensors permanently and compute whether an accident is occurring.

Activation of retention systems

The sensor signals are evaluated by intelligent algorithms, the parameters of which are optimized specifically for individual vehicles with the aid of simulations and crash tests. The airbag should not trigger from a hammer blow in the workshop, gentle impacts, bottoming out, or driving over curbstones or potholes.

The retentions systems connect to the ECU are activated when the trigger thresholds are exceeded. Current pulses activate pyrotechnic firing elements, which trigger the airbags or other retention systems.

In the event that the power supply to the ECU is interrupted by the vehicle battery as a result of the crash, a built-in power reserve provides the necessary availability.

The firing of retention systems for the relevant seat occurs only when the seat-belt buckle is connected; this is monitored by a seat-belt buckle switch.

Special functions

Advanced Rollover Sensing

The "Advanced Rollover Sensing" function uses signals from the driving-dynamics-control sensors to better detect soil-trip rollover processes. These data are used by the airbag ECU to calculate the speed vector and the lateral velocity. From these the deviation of the vehicle-motion vector from the vehicle's longitudinal axis and thereby lateral vehicle motion are determined.

Early Pole Crash Detection

The "Early Pole Crash Detection" function also uses the signals from the driving-dynamics-control sensors to better detect a side-on pole impact. In the event of a side-on pole impact the vehicle skids before the impact; this is detected with the yaw sensor. The information regarding the lateral movement of the vehicle is then used in the triggering algorithm to accelerate triggering of the side airbags.

Secondary Collision Mitigation

In the event of accidents the initial collision may be followed by further collisions by the accident vehicle, for example as a result of the driver losing control of the vehicle. This phenomenon puts both the vehicle occupants and other road users at risk. The "Secondary Collision Mitigation" function provides assistance in the event of such accidents. If an impact occurs, the airbag ECU transmits this information to the driving-dynamics control ECU. The driving-dynamics control uses specifically targeted brake interventions to slow down the vehicle or where necessary to shut down the engine. Consequential accidents can thus be avoided or their severity reduced.

Further safety measures

By networking driving-dynamics control and airbag ECU, it is possible also to utilize the detection of unstable or critical states to initiate more specifically targeted safety measures. When an unstable driving state is detected, safety measures can be introduced in stages; these measures can include closing the windows and the sliding sunroof, and tightening a reusable (reversible) motor-driven seat-belt pretensioner.

Retention systems and actuators

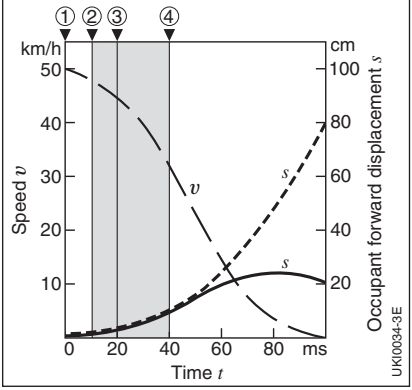
Seat belts and belt pretensioners

Seat belts

The function of seat belts is to restrain the occupants of a vehicle in their seats when the vehicle impacts against an obstacle. In this way, the occupants are already involved at an early stage in the deceleration of the vehicle during an impact (Figure 2).

Figure 2: Deceleration to standstill and forward displacement of an occupant at an impact speed of 50 km/h

① Impact,
② Firing of belt pretensioner and airbag,
③ Belt tensioned,
④ Airbag inflated.
– – – Forward displacement w/out restraint system.
—— Forward displacement w/ restraint system.



The standard type is the three-point belt with inertia-reel device. The belt buckle and the inertia-reel device are mounted on the seat or on the body (Figure 3). A loosely fitting belt (for example, when thick winter clothing is worn) results in belt slack. The softness of the clothing prevents the occupant from being involved at an early stage in the deceleration of the vehicle in the event of a collision. The occupant initially continues moving without being restrained, which reduces the protective effect of the belt. Furthermore, the slowed-down effect of the belt locking in an accident and the belt-webbing elongation contribute to belt slack.

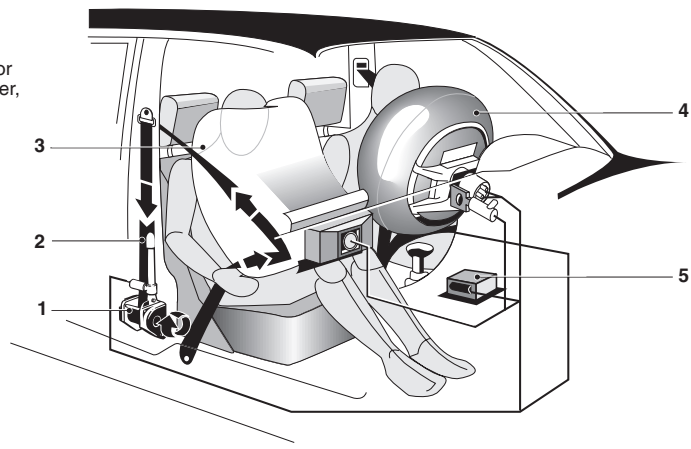
Because of belt slack three-point inertia-reel belts only afford limited protection in the event of a frontal impact at speeds in excess of 40 km/h against solid obstacles, because they cannot safely prevent the head and body from striking the steering wheel and the dashboard.

Shoulder-belt pretensioner

When activated, the shoulder-belt pretensioner eliminates the belt slack and the "film-reel effect" by rolling up and tight-

Figure 3: Occupant-protection systems with seat-belt pretensioner and front airbags

1 Inertia-reel device with seat-belt pretensioner,
2 Seat belt,
3 Front airbag for front passenger,
4 Front airbag for driver,
5 ECU.



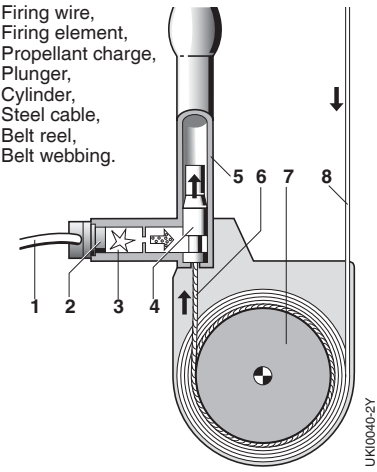
ening the belt webbing. Here, the system electrically fires a pyrotechnic propellant charge (Figure 4). The gas charge released in this process acts on a plunger, which turns the belt reel via a steel cable in such a way that it is held tightly against the occupant's body. The belt webbing is therefore tightened already before the occupant starts their forward movement. With these belt pretensioners the belt webbing can be pulled back within a period of 10 ms by up to 12 cm.

Electromechanically actuated belt pretensioners are reversible and can therefore already be activated early in a hazardous situation without there being a collision.

Buckle tightener
When triggered by a propellant charge or by spring systems, the buckle tightener pulls the seat-belt buckle back and simultaneously tightens the shoulder and lap belts. It improves the restraining effect and the protection to prevent occupants from sliding forward under the lap belt ("submarining effect").

Combination of shoulder-belt and buckle tighteners
The combination of both systems offers greater tightener travel to obtain an improved restraining effect. The two tighteners are activated for the most part simultaneously or with a slight time lag.

Figure 4: Shoulder-belt pretensioner
1 Firing wire,
2 Firing element,
3 Propellant charge,
4 Plunger,
5 Cylinder,
6 Steel cable,
7 Belt reel,
8 Belt webbing.



Belt-force limiter
The maximum force of the belt is limited in order to limit the forces on the occupant's torso in the event of serious crashes. Otherwise the occupant may suffer broken collarbones and ribs with resulting internal injuries.

This can on the one hand be performed by a mechanical belt-force limiter, which limits the belt force to around 4 kN. This is effected for example by the deformation of a torsion bar in the belt-retractor shaft or a deployment control seam in the belt.

An additional protective effect is provided by an electronically controlled belt-force limiter, which after reaching a defined forward displacement reduces the belt force by activating a firing element to 1 to 2 kN.

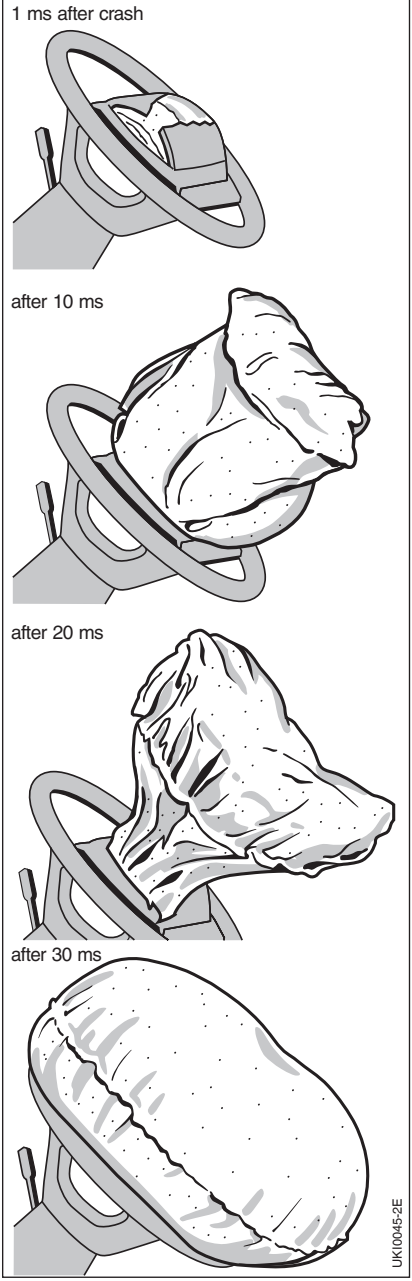
Airbags and gas inflators
Front airbags
During an accident front airbags protect the driver and the front passenger against head and chest injuries (Figure 3) which can occur when the head strikes the steering wheel or the dashboard.

To perform this function, airbags have, depending on the installation location, type of vehicle, and structure-deformation behavior of the vehicle body, different filling capacities and shapes adapted to the vehicle conditions.

Pyrotechnic gas inflators inflate the driver and passenger airbags using dynamic pyrotechnics after a vehicle impact detected by the acceleration sensors (Figure 5). In order to achieve maximum protection, the airbag must be fully inflated before the occupant comes into contact with it. When the occupant comes into contact with the airbag, the airbag is partially deflated via deflation vents. The impact energy acting on the occupant is "softly" absorbed with noncritical (in terms of injury) surface pressures and deceleration forces.

The inflation speed and the hardness of the inflated airbag can be influenced in the case of two-stage gas inflators by a delayed firing of the second stage, which

Figure 5: Highly dynamic deployment of the driver airbag



injects and additional charge of air into the airbag in the event of high crash severity.

The maximum permitted forward motion of the driver until the airbag on the driver's side has inflated is approx. 12.5 cm. This corresponds to a time of approximately 40 ms in the case of an impact at 50 km/h on a hard obstacle. 10 ms after the impact the gas inflator is activated and after 30 ms the airbag is inflated. The airbag deflates after another 80 to 100 ms through the deflation vents.

Depowered airbags

So as not to put children and very light adults on the front passenger seat at risk from the inflation process, "depowered airbags" with a reduced inflation force are used above all in the USA.

To this end, in less serious accidents at reduced speed up to around 30 km/h the airbags are inflated with 20 to 30 % less gas-inflator output, reducing the inflation speed and the hardness of the inflated airbag. This reduces the risk of injury for occupants who are "out of position" (posture diverging from the normal seating position).

With the low-risk deployment method, depending on the severity of the impact, only the first front-airbag stage is fired or else both inflator stages to bring the full gas-inflator output to bear.

Airbags with active ventilation system

These airbags have a controllable deflation valve to maintain the internal pressure of the airbag constant even if an occupant falls against it and to minimize occupant trauma. A simpler version is an airbag with "intelligent vents". These vents remain closed (so that the airbag does not deflate) until the pressure increase resulting from the impact of the occupant causes them to open and allow the airbag to deflate. As a result, the airbag's energy absorption capacity is fully maintained until the point at which its motion-damping function comes into effect.

Knee airbag

In a few vehicle types, front airbags also operate in conjunction with inflatable knee pads, which safeguard the "ride down benefit", i.e., the speed decrease of the

occupants together with the speed decrease of the passenger cell. This ensures that the upper body and head describe the rotational forward motion which is needed for the airbag to provide optimum protection. Furthermore, the knee airbag prevents contact with the dashboard support and thereby reduces the risk of injury in this area.

Side airbags
Side airbags which inflate along the length of the roof lining for head protection (e.g., inflatable tubular system, window bag, inflatable curtain) or for upper-body protection from the door or seat backrest (thorax bag) are designed in the event of a side impact to push the occupants away from the impact location and facilitate a cushioning of the impact and thereby protect them from injury.

The challenge for side airbags is the shorter time available for inflation compared with front airbags. Here there is no vehicle crumple zone and the spacing between the occupant and the vehicle body is very small. The time for crash sensing and activation of the side airbags must therefore be approximately 5 to 10 ms in the case of severe side impacts. The inflation duration for thorax bags may not exceed 10 ms.

Additional airbag equipment
A further improvement in the restraining effect is optionally provided in individual vehicles by airbags integrated in the thorax section of the seat belt (air belts, inflatable tubular torso restraints or bag-in-belt systems), which reduce the risk of broken ribs.

Inflatable headrests protect against whiplash and cervical-vertebrae injuries in the event of a rear-end impact (adaptive headrests), inflatable carpets on the other hand protect against foot and ankle injuries. Active seats, in which an airbag in the front section of the seat surface is inflated, impede the occupant's forward sliding movement (submarining effect).

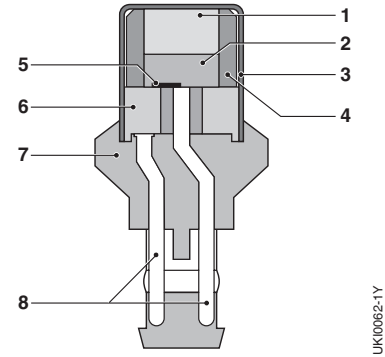
Gas inflators
The gas inflator fills the airbag with gas and actuates the belt pretensioner. The pyrotechnic gas inflator contains a firing pellet, which when activated in turn ignites a solid propellant. The gas passes through a metal filter and the inflator's outlet ducts into the airbag.

The firing pellet (Figure 6) contains a reservoir holding the propellant charge and a priming wire. The firing pellet is connected to the airbag ECU via the connector pins and a two-wire circuit. To trigger the airbag, the ECU uses firing output stages to generate a current, which flows inside the firing pellet through the priming wire. This wire glows and activates the propellant charge.

The driver airbag fitted in the steering-wheel hub (volume approximately 60 liters) and the passenger airbag fitted in the glovebox space (approximately 120 liters) are inflated approximately 30 ms after firing.

Enhanced protective functions
Rollover protection
If a car turns over, there is the danger that non-belted occupants may be thrown through the side windows, or that parts of the bodies of belted occupants (e.g., arms) may protrude from the vehicle and

Figure 6: Firing pellet
1 Propellant charge,
2 Detonator,
3 Cap,
4 Charge holder,
5 Priming wire,
6 Firing head,
7 Housing,
8 Connector pins.



incur serious injury. To provide protection in such cases, existing restraint systems such as seat-belt pretensioners and side and head airbags are activated. In convertibles, the extendable rollover bars or the extendable head restraints are triggered.

A rollover is sensed in the vehicle transverse and vertical directions (y- and z-axes) with micromechanical yaw sensors and high-resolution acceleration sensors integrated in the airbag ECU. The yaw sensor is the main sensor, while the y- and z-axis acceleration sensors are used both to check plausibility and to detect the type of rollover (slope, gradient, curb-impact or soil-trip rollover).

Depending on the rollover situation, yaw rate and lateral acceleration, the retention systems and actuators are activated and deployed after 30 to 3000 ms.

Pedestrian protection
This function is intended to avoid injuries as far as possible in the event of vehicles colliding with pedestrians. Leg and head injuries can be reduced by appropriate designs of bumpers and engine hood. In addition, in a crash situation the hood can be lifted or an external airbag can cushion the impact (Figure 7). The sensor technology for this purpose (acceleration sensors or a tube with pressure sensors, see Figure 1) is integrated in the bumper. Appropriate tests are an integral part for example of the Euro NCAP crash test.

Occupant sensing

There are various options on the market for occupant classification.

iBolt
Occupant classification can be performed with the iBolt (“intelligent” bolt), an absolute-weight-measuring method. These iBolts (see Figure 1) measure the forces involved and secure the seat frame (suspended seat) on the sliding base, replacing the four mounting screws that are otherwise used. They measure the weight-dependent change in the gap between the bolt sleeve and the internal bolt with integral Hall-element IC connected to the sliding base (Figure 8).

Seat-occupancy mats
Seat-occupancy mats combined with pressure-sensitive elements and matching sensor also perform this function (Figure 9). The information is sent to the ECU and incorporated in the triggering strategy for the means of protection. The seats can be adjusted when backrest angle and position are sensed with suitable actuators during the crash to better protect the occupants.

Deactivation of front-passenger airbag
A manually actuated deactivation switch can be used to disable the passenger airbag. This is required if certain child seats are used in this space.

Figure 7: Raised engine hood and external airbag as pedestrian protection

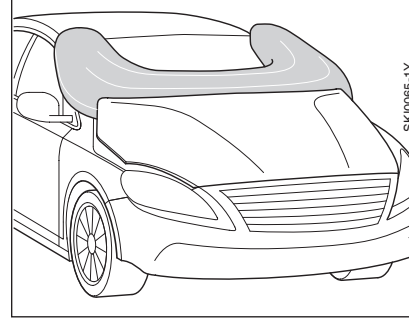
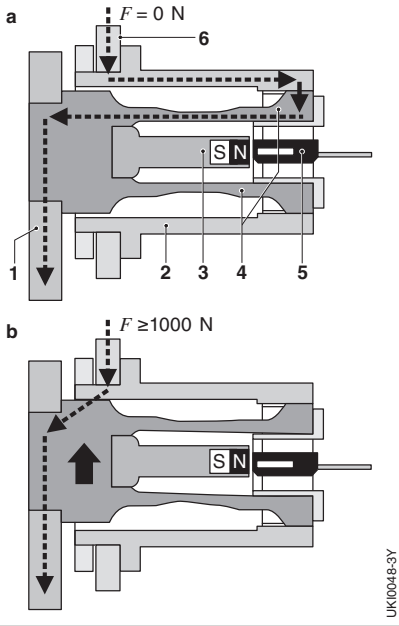
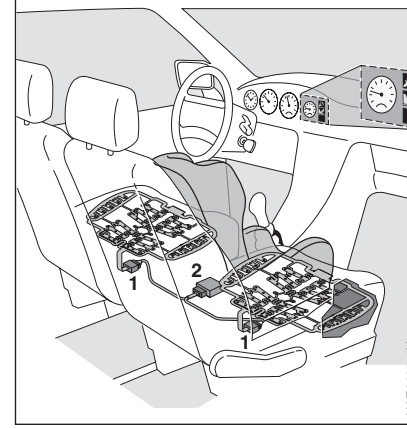


Figure 8: Force-measuring iBolt
a) Rest position,
b) In operation, i.e., in overload stop.
1 Sliding base,
2 Sleeve,
3 Magnet holder,
4 Double bending beam (spring),
5 Hall IC,
6 Seat frame.



There is an increasing availability of standardized anchoring systems (ISOFIX child seats). Switches integrated in the anchoring locks initiate an automatic front-passenger airbag deactivation, which must be indicated in the instrument cluster.

Figure 9: Installation position of the OC sensor mats in the vehicle front seats
1 OC ECU,
2 Airbag ECU.



Out-of-position detection

In order to prevent injuries being caused by airbags to occupants who are “out of position” (e.g., are leaning far forwards) or to small children in reboard child seats (seat position pointing towards the rear), the triggering of the front airbag and its inflation must be adapted to the situation.

Enhanced detection of the occupant position and adaptation of the retention systems and their activation are important for future automated driving, since the challenges faced by passive safety increase with the freedom of movement gained by the occupants.

Integral-safety systems

Overview

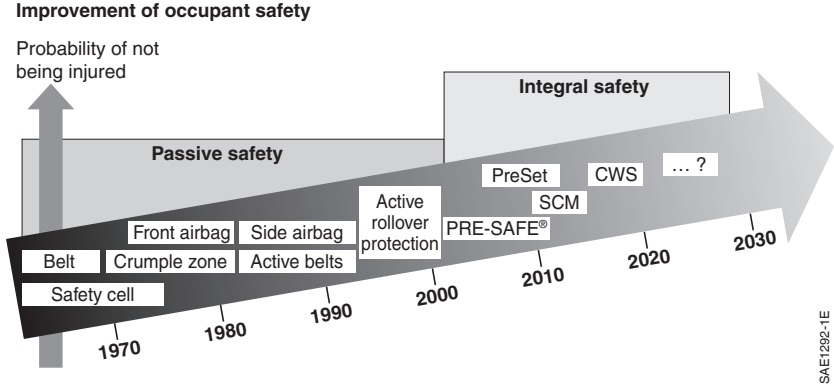
The road to automated driving will be accompanied in the coming years by a marked increase in the market penetration of environment-sensor systems. The development of new assistance and driving functions necessitated by the introduction of automated driving leads to the availability of additional information on vehicle, environment, and occupants. This extended and networked data basis permits the development of new integral safety concepts with the aim of adapting the vehicle's protective measures to the respective accident situation in the event of unavoidable accidents. Predictive situation analysis in particular facilitates safety measures for preventing accidents or lessening the accident severity. The networking of active and passive safety measures creates the precondition for situational adaptation of the vehicle's automatic protective responses to the respective accident events.

Motivation

It has been possible in the past few years through numerous measures in the automotive industry, traffic research and legislation to reduce the number of traffic fatalities and serious injuries. The fitting prescribed by law of all new passenger cars with driving-dynamics control (Electronic Stability Program, ESP®; Electronic Stability Control, ESC), the introduction of antilock braking systems (ABS) for motorcycles, and lane- change warning systems are mentioned by way of example. Numerous serious accidents still occur despite these measures [1].

Accident protection does not lose in importance even when it comes to automated driving. There is a need for action, since integral safety forms the fallback level of automated driving. The integral-safety approach offers new solutions which will both take into account the individuality of the individual occupant and raise the traffic-safety level through the networking of sensor and safety systems.

Figure 1: Market introduction of integral-safety functions (source: Bosch)
PRE-SAFE®: Activation e.g. of reversible belt pretensioners and seat.
PreSet: Improved accident-severity determination by means of environment-sensor technology.
SCM: Secondary Collision Mitigation; braking of vehicle after primary collision to avoid secondary collision.
CWS: Collision Warning System; activation of reversible belt pretensioners when leaving the road in the direction of an extreme surface.



Technology of the present day
Integral safety is based on the networking of components and algorithms of various safety and comfort and convenience systems in the vehicle (e.g., passive-safety retention systems, driving-dynamics control, driver-assistance systems, passenger-compartment sensing). Here, the entire chain of events of an unavoidable accident before, during and after the collision is analyzed to minimize the risk of injury to vehicle occupants and other road users.

The first integral-safety systems were launched in 2002 by Mercedes-Benz under the name PRE-SAFE® in the S-Class. Since 2009 the front close-range radar sensors of the driver-assistance package have also been used with this system to detect potential collision objects with preventive reversible seat-belt tightening. In 2013 this principle was carried over to the vehicle rear end.

A further step in the direction of integral safety was made by the "PRE-SAFE impulse side" function introduced in 2016 for the front seats, in which for the first time pyrotechnics are activated exclusively on the basis of environment-sensor technology. Side-mounted radar-sensor technology detects an unavoidable side impact, whereupon an actuator integrated in the seat imparts to the occupant an impulse towards the center of the vehicle [2].

Figure 1 shows the timeline of the transition from classic passive safety to integral safety. Safety functions and safety components already introduced in the market are depicted.

Solution approaches

Challenges
Since the introduction of radar technology for driver-assistance functions like Adaptive Cruise Control (ACC) or Automatic Emergency Braking (AEB) in the automobile sector, new function ideas have arisen which are founded on the use of environment data to better estimate the accident severity (e.g., the PreSet function for determining accident severity). However, such functions have not been widely used due to the low equipment rate with Adaptive Cruise Control. This situation is changing, as Automatic Emergency Braking has since 2014 been rewarded in the Euro NCAP protocol. As a result, the proportion (equipment rate) of new cars fitted with radar sensors has increased. In spite of this conducive boundary condition there are hurdles for the market introduction of new integral-safety systems:

- Effects of new safety functions only become visible in the accident statistics after a delay, as the rising equipment rate in new-vehicle registrations is set against an old vehicle population.
- Dedicated environment sensors and processing units only for the integral-safety functions have not gained acceptance in the past for cost reasons, which hinders the introduction of independent integral-safety systems.
- Networked functions with numerous data interfaces between assistance and restraint systems entail additional application and validation outlay.

The penetration times of integral-safety systems will follow the market developments with regard to driver assistance, active safety, and automated driving. In the course of the introduction of automatic driving functions the equipment rate with environment and passenger-compartment sensors and with domain ECUs will increase [3]. The introduction of automatic driving functions will ensure the market introduction and market penetration of environment and passenger-compartment sensors and of appropriate domain ECUs. At the same time, occupant protection faces hitherto unknown challenges as a result of these functions. Possible new seating positions,

shifting of the occupant positions during automatic braking or steering operations or larger objects between occupants and airbags change the boundary conditions for the optimum use of retention systems in the event of a collision risk.

Even fundamental elements of present-day vehicles, such as function and E/E architecture, undergo a development process in the course of developments in the fields of electromobility, automated driving, and connectivity. New requirements with regard to safety, communication speed, computing outlay, latencies, and costs drive this development process, which is also useful in the introduction of integral-safety systems.

Application examples

Accident types with head-on collision

Head-on-accident scenarios as depicted in Figure 2 highlight an important sphere of activity for integral safety. The frequently occurring rear-end collision (Figure 2a) can already today be firmly detected in terms of its development by the use of environment-sensor technology. The growing number of vehicles which are equipped with an automatic emergency brake system will reduce the frequency and severity of this type of accident.

Intersection accidents with side collisions as pictured in Figure 2e constitute a further area of use for integral-safety functions. Injury severity is frequently in the middle range. Intersection-assistance systems with left turn-off assist and cross-traffic assist will greatly reduce the frequency of such accidents (see Driver-assistance systems).

The introduction of Electronic Stability Programs, lane-departure warning systems, and functions for detecting driver fatigue brought about a reduction of in collisions caused by the vehicle leaving the road (e.g., tree collision, Figure 2d). This accident type is nevertheless still very relevant today on account of the injury severity often associated with it.

Accidents involving oncoming vehicles exhibit a high injury probability (Figure 2c). Active-safety and automation measures prove to be the least effective in these accident scenarios. Integral safety is therefore focused on this accident type.

There is need for improvement in the case of accident types which are not reproduced by laboratory tests or for which conventional contact-sensor technology exhibits limitations. These conditions apply for example to driving under a truck

Figure 2: Relevant application cases in the area of head-on collisions (source: Bosch)
a) Rear-end collision,
b) Driving under a truck,
c) Head-on collision,
d) Tree collision,
e) Side collision.



(Figure 2b) or a central impact with a tree with a small overlap (Figure 2d). In both cases the sensors installed in the supporting frame respond only at a relatively late stage after the body has undergone marked deformation.

Function examples
The risk of injury increases if the moment of airbag deployment is not optimally matched to the forward displacement of the occupant. This situation ensues when the airbag is fired too late or when the occupant is too close to the airbag due to forward displacement caused by a braking operation ([4], [5]). Reference is made by way of example to two functions which address this problem: Integrated Collision Detection Front (IDF) and Pre-Crash Positioning (PCP).

Integrated Collision Detection Front (IDF)
The IDF function uses environment-sensor data (radar, video) in conjunction with conventional contact-sensor technology to trigger belt pretensioners and airbags at an earlier stage. Aside from an accurate estimation of the relative impact speed and the moment of impact, the merged sensor system offers the following advantages in terms of optimizing the efficiency of the retention systems:
– more precise description of the collision object,
– optimization of the interfaces between ADAS (Advanced Driver Assistance Systems) and restraint system.

Pre-Crash Positioning (PCP)
The PCP function coordinates in scenarios depicted in Figure 2 automated emergency braking with the restraint system. The chronological sequences of forces acting on the occupants, braking operations, and activations of the retention systems must be matched in such a way as to minimize the load on the occupants. A first instance of this function serves to position the occupants at the right time in the phase prior to the collision by means of reversible seatbelt tightening.

Further developments
Later versions of the IDF and PCP functions will, along similar lines to the further development of the automatic emergency brake systems, take into account more complex relationships. The first expansion stage handles the rear-end impact using front-sensor technology with the merging of radar and video. The same sensor configuration is used in the case of oncoming vehicles. However, this accident type requires the development of more complex decision algorithms and additional safeguarding effort to avoid bad decisions. Expansion to include intersection accidents necessitates the additional installation of Mid-Range-Radar (MRR) with a very large side beam width.

Individual safety
On account of legal regulations and consumer tests today's passive-safety systems are optimized for certain occupant heights and weights. Adaptive retention systems combined with suitable activation offer the potential for raising the safety level also for occupants that deviate from the standard sizes used in the tests. One possible solution approach uses occupant-specific variables such as age, height, weight or gender for the purpose of individually activating existing and future adaptive retention systems.

These data are acquired via passenger-compartment-sensor technology, supplemented by user inputs and plausibility checks. For example, the user enters once via his smart phone or an HMI (Human Machine Interface) integrated in the vehicle his personal data, which are linked with an image recorded by him. On each new entry into the vehicle a unique identification of the occupant is performed by means of passenger-compartment camera and image-processing algorithms with corresponding parameterization of the retention systems. If the occupant is not identified, the usual high safety level is still available in the default settings.

In view of the anticipated increase in the use of automatic and semi-automatic driving functions, an increase in the number of passenger-compartment cameras being installed is also to be expected. Personal optimization of the safety functions therefore does not require additional sensors. It is possible to fall back on passenger-compartment-sensor technology to be introduced to safeguard automated driving.

Safeguarding

A major challenge faced by integral-safety functions is environment and passenger-compartment sensing and correct situation assessment. As well as verifying correct functionality, particular attention is placed on safeguarding against unwanted false triggering of the retention systems. A typical situation that is to be avoided is the erroneous reporting of a collision object without such an object actually existing.

A further scenario to be avoided is as follows: An object is correctly detected, however an erroneous trajectory computation suggests a collision course and activates retention systems although the object is actually driving past the system vehicle. Falling into the same category are driving situations in which an object on a collision course is correctly detected, but after retention systems have been triggered the collision is not avoided [2].

In spite of data fusion and the associated redundancy, there are complex situations whose detection and correct interpretation must be verified. To guarantee this, the testing of integral-safety systems must record as wide a variety of real traffic situations as possible.

Testing of integral-safety system according to today's methodology exceeds both the cost and the time framework for release in the automotive industry. A first step towards reducing expenditure while simultaneously guaranteeing safety is to break down the wide range of real situations into a limited number of simulated test-track or laboratory situations which are representative of reality.

A supplementing approach is to use virtual instead of real test environments. Virtual test environments permit the evaluation of safety functions in the early development phase and thus fast prototype construction. Furthermore, they enable safeguarding of the functions to be developed in complex traffic situations that are hard to reproduce in real driving or test operation.

Outlook

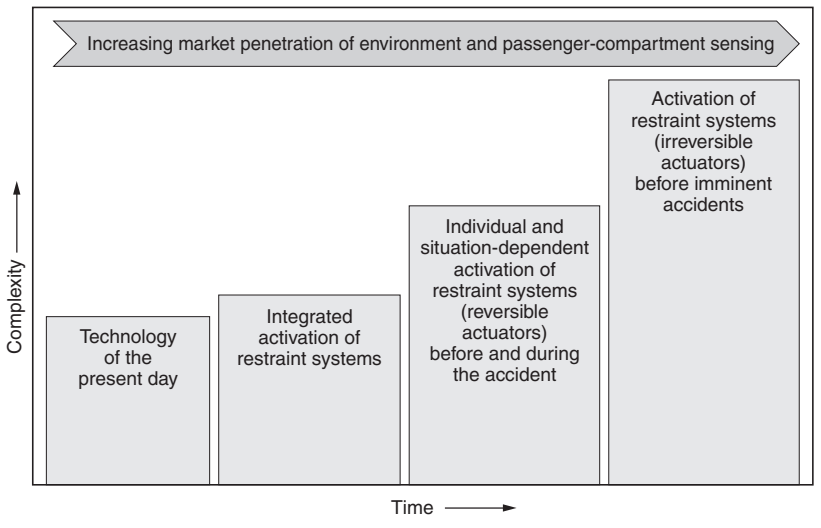
Starting out from conventional vehicle safety with deployment of the airbags after an impact has been detected, the trend is towards adaptation of the trigger thresholds of retention systems on the basis of networked object data by environment sensors, passenger-compartment sensors, and driving-dynamics information. In this expansion stage of the integral-safety systems only reversible protection systems will be activated prior to the collision.

Functions with activation of irreversible actuators exclusively on the basis of data from environment sensors will constitute the next development stage of integral safety (Figure 3).

References

[1] T. Lich et al.: "Is there a broken trend in traffic safety in Germany? Model based approach describing the relation between traffic fatalities in Germany and environmental conditions", ESAR Conference, Hanover, Germany, 2014.
[2] J. Richert, R. Bogenrieder, U. Merz, R. Schöneburg: "PRE-SAFE® Impuls Seite – Vorauslösendes Rückhaltsystem bei drohendem Seitenaufprall – Chance für den Insassenschutz, Herausforderung der Umfeldsensorik", 10th VDI Conference Vehicle Safety Safety 2.0, Berlin, 2015.
[3] J. Becker, S. Kammel, O. Pink, M. Fausten: "Bosch's approach toward automated driving". at – Automatisierungstechnik 2015; 63(3): 180 – 190.
[4] G. Gstrein, W. Sinz, W. Eberle, J. Richert, W. Bullinger: "Improvement of airbag performance through pre-triggering". Paper Number 09 – 0229, Proceedings of 18th Enhanced Safety of Vehicles, ESV 21, Stuttgart, Germany, 2009.
[5] H. Freienstein, T. Engelberg, H. Bothe, R. Watts: "3-D Video Sensor for Dynamic Out-of-Position Sensing, Occupant Classification and Additional Sensor Functions". Paper No. 2005-01-1232, SAE World Congress 2005, Detroit, USA.

Figure 3: Development trend of systems for integral safety (source: Bosch)



Locking systems

Task and design

The locking system of a vehicle includes all those components that facilitate authorized access to and exit from the vehicle. These components also prevent the inadvertent opening of doors, gates, and lids and unauthorized entry. The tasks they must perform are as follows:

– facilitating vehicle access (e.g., via the side doors) to all authorized persons and access to the interior (e.g., via engine hood, luggage compartment, glove compartment, or fuel filler cap),
– facilitating the opening of doors from the inside and outside,
– protecting occupants against inadvertent door opening,
– keeping doors, gates, and lids securely closed during vehicle operation and in crash situations and immobilizing,
– protecting the vehicle against theft when the vehicle is parked,
– detecting an opened or closed door.

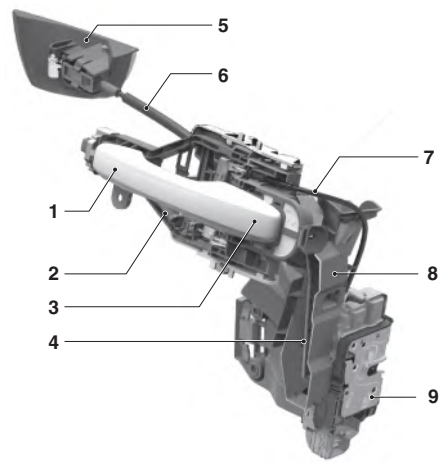
The locking system can consist of the following components (Figure 1):

– identification system (mechanical key, remote key, Near Field Communication via chip in the remote key or smart phone),
– outer control (mechanical pull or flipper-type handle, pushbutton or proximity-type sensor),
– outer locking (mechanical lock barrel, pushbutton or proximity-type sensor),
– inner locking (mechanical pull handle, pushbutton or proximity-type sensor),
– inner locking (mechanical control, pushbutton or proximity-type sensor),
– lock (mechanical, mechanical-electric, electric),
– lock holder,
– connecting elements (rods, Bowden cables, electrical leads),
– door, gate, and lid drives,
– ECU (door ECU or central ECU).

The further types refer to the locking system in the side door, since it combines the most functions and the greatest complexity.

Figure 1: Components of the locking system in a vehicle side door

1 Outer handle,
2 Outer-handle plate,
3 Lock barrel
(behind the cover plate),
4 Outer-locking rod,
5 Inner handle,
6 Bowden cable for inner handle,
7 Bowden cable for outer handle,
8 Retaining bracket,
9 Side-door lock.



UKT0091-1Y

Access authorization

Standard
Today's standard is represented by remote keys, which at the touch of a button by the user send a coded data packet at a UHF frequency (433 MHz or 868 MHz) to the vehicle. This data packet changes with each actuation in order to prevent “eavesdropping” (see Immobilizer). These data packets typically consists of a fixed-code part and a rolling-code part which are generated according to an algorithm. The receiver decrypts this code with the same algorithm and thereby identifies permissible combinations. Each code can only be used once. Once the received data packet is identified as permissible, the locks are electrically unlocked and the vehicle owner can open the doors by pulling on the outer control. The mechanical key now only acts as an emergency actuator in the event of a remote-key malfunction or a power failure.

Keyless entry
So-called passive-entry or keyless-entry access systems are becoming increasingly popular. These do not require any electrical-unlocking action by the vehicle owner. It is enough for the owner to be carrying the ID transmitter about their person. A chip inserted in the remote key or a suitable smart phone can serve as the ID transmitter (Near Field Communication). With these systems the opening request is automatically detected by the vehicle – for instance when approaching a capacitive sensor in the outer door handle, when pulling on the outer door handle, or when operating a pushbutton in the handle. The central ECU in the vehicle sends a signal at a low frequency (125 kHz with a range of around one meter) to the chip in the ID transmitter. The latter reacts to a UHF frequency with a range of up to 20 meters and sends back a coded reply sequence to the vehicle. Access is authorized and the doors are unlocked.

The opening speed is crucial to the user's comfort and convenience. Both the response time of the identification protocol and the unlocking of the doors must be so short that there is no waiting time. The user should not have to pull repeatedly on the outer door handle before the door opens.

In a passive-entry system the reply sequence also changes with each request so as to prevent “eavesdropping”. However, a security risk is posed by so-called “relay-station attacks”. In such attacks the range of the signals of vehicle and ID transmitter are extended with portable transmit/receive stations. One of these devices must be close to the vehicle to be able to receive and boost the low-frequency signal. In this way, the ID transmitter can also be scanned unnoticed by the user over large distances and the vehicle unlocked or started. In newer systems the passive-entry function can be deactivated to prevent such manipulations.

Lock design

Locking mechanism and lock holder

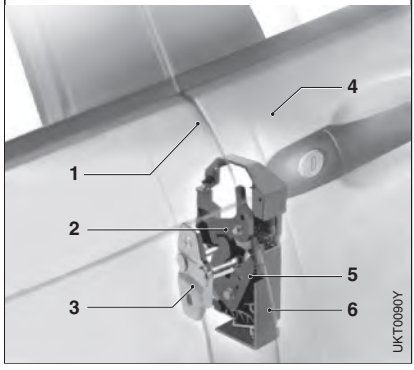
Interaction

The locking mechanism (Figure 2) in the lock establishes a separable connection with the lock holder mounted on the body. Indirectly locking systems have gained acceptance on the market. The holding-closed and locking functions are split between two components – the spagnolet (holding closed) and the locking pawl (locking). The lock holder is enclosed by the spagnolet when the door is closed. The spagnolet is driven by the lock holder, rotated, and locked in the door opening direction by the spring-loaded locking pawl. These indirect systems require only minimal opening forces.

During the closing operation the locking mechanism must be able to compensate door tolerances in relation to the body. The position of the door in the closing direction is determined by the lock holder. This is adjusted once during vehicle assembly in order to set the door alignment and the gap dimensions.

Figure 2: Sectional view of a typical locking mechanism with lock holder

1 B-pillar,
2 Locking pawl (in the door),
3 Lock holder (in the B-pillar),
4 Front passenger door,
5 Spagnolet,
6 Lock housing.



Preliminary detent and primary detent

Statutory requirements such as two-detent capability, tensile values, and 30-*g* strength must be satisfied when it comes to locking-mechanism design. Two-detent capability means that the spagnolet can be held in two positions by the locking pawl – the preliminary and the primary detent. In the primary detent the door is in the desired closing position, i.e., flush with the body, and the door seals are fully compressed. The preliminary detent is a position roughly 5 to 7 mm before the primary detent; in this case the door is not fully closed and the door seals are only partly compressed. In this position watertightness is not guaranteed and driving noises can be clearly heard.

The preliminary detent fulfills two main functions: If the lock inadvertently leaves the primary-detent position, the door is caught again and does not swing open. If the swing is too small when closing, the door remains at least in the preliminary detent and cannot open again inadvertently. Premium locks with an electric closing aid go a step further: They move the door with the aid of an additional motor automatically from the preliminary to the primary detent, slamming the door shut is no longer necessary.

Opening chain

To open the door, the locking of the spagnolet by the locking pawl must be removed. The system here is strained by the pressure of the compressed door seals. The locking pawl is connected with the door handles via a lever chain of connecting elements such as rods or Bowden cables. This is also known as the opening chain, which transfers the handle travel up to the locking pawl. The action of swinging out the locking pawl disengages the tension generated by the seal pressure, the door is released, and the typical opening sound can be heard.

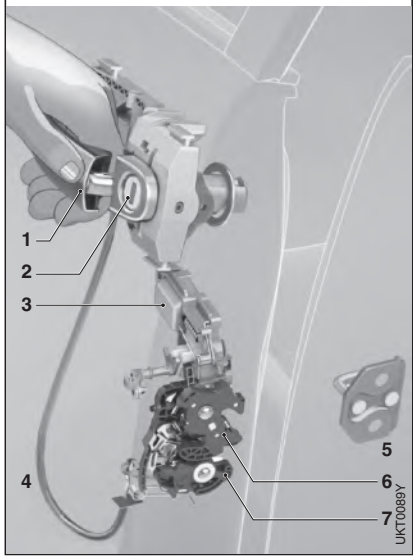
Aside from the pull (Figure 3) and flipper-type variants, the outer handle can also feature pushbuttons or proximity-type sensors. This provides solutions in the lock in which a motor lifts out the locking pawl and thereby opens the lock. This is already standard in rear locks.

To control entry and exit one or more switchable couplings are required in the

lock which interrupt the opening chain for the outer or inner handle. The door is then locked.

Outer locking
Most of today's vehicles have a central-locking system, which is activated from the outside via the remote key. If the vehicle is without electric power and hence the central-locking system is no longer working, a mechanical lock barrel usually mounted on the driver's door in the outer handle (often under a protective cap) enables the vehicle to be manually locked and unlocked. This is also the case for vehicles with mechanical locks, i.e., without central locking. Vehicles with the passive-entry function as a rule incorporate another pushbutton or a sensor surface in the outer handle which must be actuated to lock the vehicle.

Figure 3: Opening chain for the outer control
1 Outer door handle, 2 Lock barrel,
3 Central-locking motor,
4 Bowden cable for outer handle,
5 Lock holder in the B-pillar (striker),
6 Spagnolet, 7 Locking pawl.



Inner locking
The inner locking, the “knob” in the inside door panel, has already been dropped by some manufacturers. It is replaced in part by an LED in the driver's door in combination with an electric pushbutton in the dashboard or the door and an emergency-locking facility on the lock. This emergency-locking facility enables the vehicle to be locked even when it is without electric power and protected against theft if the battery is flat. The locking action is then performed by means of a lever on the front of the door; this must be opened for operation purposes. Sensor surfaces on the inside of the door or gesture control to activate the central locking from the inside are also conceivable future developments.

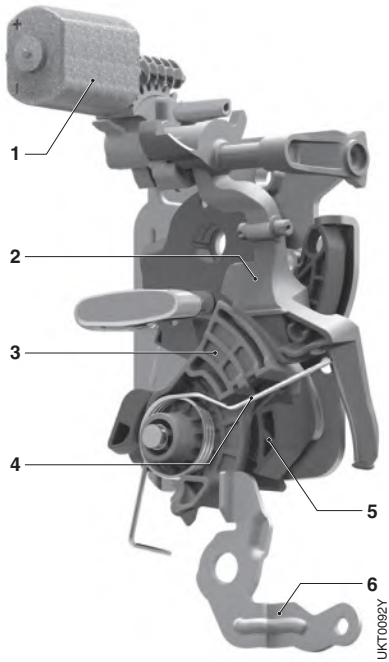
Central locking via DC motor
The couplings can be switched either manually by the outer and inner locking or by a DC motor incorporated in the lock (Figure 4). The latter is the case in a central-locking system. In this case all four doors and the rear lock are electrically locked simultaneously. The locking function can be manually canceled by pulling the inside door handle once or twice. The electric central locking is activated via the radio remote control and a passenger-compartment button, for example in the dashboard.

Child-lock function
The child-lock function for the rear doors is a special feature. This function merely disengages the inner handle, preventing rear occupants from being able to open the rear doors from the inside. Depending on the lock used, the door can however be unlocked (but not opened) from the inside, enabling persons outside the locked vehicle to gain entry. The child lock can be activated manually or by a motor in the lock. This is done manually by means of a lever which can be accessed only when the door is open.

Requirements

A side-door lock is subject to an extensive range of requirements. Aside from the legal requirements FMVSS 206 [1] and ECE R11 [2], each manufacturer specifies its own statement of requirements which extends far beyond the statutory provisions. Table 1 provides an overview of the essential requirements relating to a side-door lock as typically apply today.

Figure 4: Typical couplings in the outer and inner opening chain of a side-door lock
1 Motor for central locking,
2 Locking lever,
3 Release lever,
4 Coupling spring
(unlocked in pictured position),
5 Locking pawl,
6 Outer-control lever.



Lock layout

The primary concern when it comes to the layout and design of the lock as a safety component is functional robustness. This is followed by criteria such as haptics and acoustics. This involves a pleasant operating feel and a closing and actuator noise that is to impart a high-quality look and feel. Depending on the functional equipment, a side-door lock can consist of 30 to 100 individual parts which interact with each other.

The tensile values or tensile strength of a locking system describe the forces that the locking system must withstand without disengaging. Resilient elements – e.g., elastomer buffers – are required in the lock to prevent door movement and associated noises during vehicle operation. This ensures good closability on the part of the door and the best possible immobilization. The buffers serve to reduce the energy of the swing motion when the door is closed.

During vehicle operation relative movements between the door and the body and consequently also between the lock holder and the lock occur. These can cause noises such as cracking or creaking on account of stick-slip effects particularly in the event of contamination by dust. Material pairings, contact geometries, the rigidity of the lock, and protection against dust accumulation are of crucial importance to the design of the lock.

Market segments, side-door locks

The type of lock that has gained acceptance today has an integrated design in which the mechanical, electromechanical, and electrical components are integrated in a housing – the so-called system lock.

The market for side-door locks is divided into three fundamental segments:

– simple locks with electric central locking (Figure 5a),
– volume locks with electric central locking, theft deterrence, and child lock (Figure 5b),
– premium locks with electric central locking, theft deterrence, electric child lock, electric opening and pulling closed (Figure 5c).

Manufacture
A lock is made up of stampings, injection-molded plastic parts, cold-upsetting parts, springs, sensors, and DC motors. They are manufactured using a variety of production techniques such as riveting, screwing, gluing, and laser welding.

Functional robustness
Functional robustness is verified as standard by means of test catalogs harmonized with the vehicle manufacturer. These are based on the requirements set out in Table 1 and are conducted on test equipment in laboratories. The vehicle manufacturer often conducts its own tests in vehicles as well.

Validation
Real tests cannot take into account all the influencing factors that occur during series production over a number of years and millions of products. These include for example geometric tolerances of individual parts or deviations from material characteristic values. This can only be delivered by simulation calculations since many factors can be taken into account here with justifiable expenditure.
Based on the typical requirements as set out in Table 1 the following are used for the purpose of design and layout:
– virtual strength and structure tests using the finite-element method (FEM, Figure 6),

Table 1: Typical requirements with regard to a side-door lock

Temperature range	– Continuous operation: –40 °C to +80 °C – Refinishing: 120 °C, 1 hour – Extreme load: 300 °C, 2 minutes
Number of door closings	– 100,000 cycles
Number of electric lockings (central locking)	– 60,000 cycles
Tensile forces	– 9 kN in primary detent longitudinal acc. to FMVSS 206 – 8 kN in primary detent transverse acc. to FMVSS 206 – Requirement of the original-equipment manufacturers (OEM) is today double FMVSS 206 and higher
Environmental requirements	– Functional robustness against water, dust and various liquids in all temperature ranges including cyclic repetition with freezing – Corrosion requirements: 144 to 720 h in salt spray
Control forces and control travel at the lock	– As a rule specification of the control forces before and after continuous running for a defined door-seal pressure (100 to 300 N) and control speed at the lock (100 to 300 mm/min); typically 15 to 20 N – Opening capability after an accident up to 5000 N tension
Acoustics	– No rattling or creaking noises – Pleasant closing and actuator noises
Push-in forces	– Force to press the lock holder into the primary-detent position: as a rule under 50 N
Electrical requirements	– Operating voltage 9 to 16 V, start assistance 24 V – Max. permissible current ≈ 5 to 8 A – Control times for central locking, theft deterrence, electric child lock 50 to 500 ms – Actuating forces for central locking with inner-locking knob: 8 N – Cut-off-voltage peaks: –75 V to +150 V – Electromagnetic compatibility as per EMV3 acc. to CISPR 25 [3]
Acceleration resistance	– Centrifuge test with 30 g stationary acceleration – Sled test with acceleration profile
Extreme requirements (misuse)	– Under other drop tests, up to 10 to 15 times increased control forces, up to 1000 mm/s increased control speeds

– virtual functional tests based on multiple-body simulation (MBS),
– statistical tolerance calculations based on the Monte Carlo method.

Simulation calculations are also carried out for the progression of the control forces in order to achieve good lock haptics. The effectiveness of design optimizations can thus be easily verified. In the case depicted in Figure 7 it was possible to reduce the control force from 45.8 N to 36.6 N with the lever ratios and contact points being optimized.

The control-travel figures of the lock with tolerances are crucial to the functionality of the locking system. Here many individual spatial tolerances interact in the control chain and these are superimposed with the aid of statistical calculations, the so-called Monte Carlo method. The result is control-travel figures with the associated tolerances and the probability of their occurrence.

Acoustic simulations are also possible and replace complex measurements at least in part. Optimizations can thus already be defined in the design phase, significantly reducing complex prototypes and measurements. New software tools today also facilitate flow simulations for the assessment and optimization of dust and water robustness.

In spite of the ever-increasing accuracy of simulation methods and material models, real test still cannot be completely replaced. This applies above all to environmental influences such as dust, water, and freezing and also to specific acoustic

Figure 6: Comparison of real and simulated tensile test using the finite-element method
a) Real depiction of the lock,
b) Calculation with finite-element method.

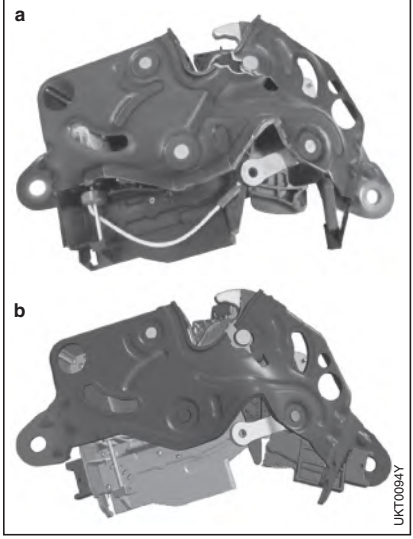
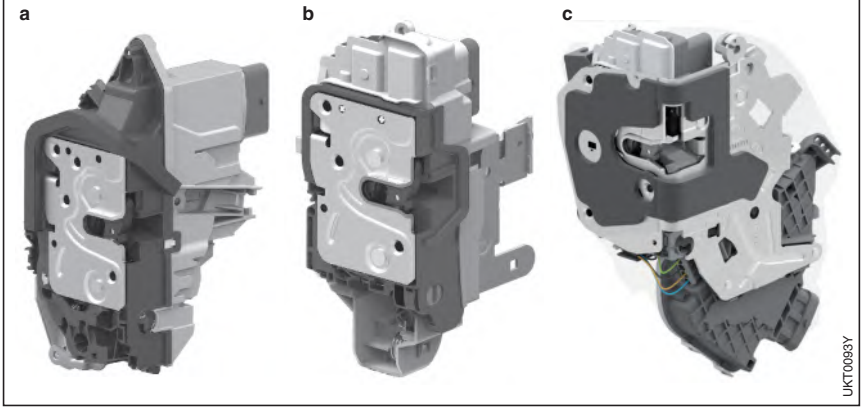


Figure 5: Lock types
a) Simple lock,
b) Volume lock,
c) Premium lock.



effects such as creaking. Realistic laboratory tests are carried out to avoid costly, complex vehicles tests with test drives.

To obtain meaningful tests in the laboratory, the dust must conform to realistic conditions with regard to composition and particle size. The same applies to the density of the dust and the air flow, which must reproduce the conditions in the door as accurately as possible. A dust-accumulation picture as from real test procedures can then be generated in the laboratory with a dust chamber.

Dust and other environmental influences such as dirt also play a role in the previously mentioned phenomenon of creaking. This arises due to friction between the door and the body in the lock – or more exactly by the lock components which are in contact with the lock holder. The occurrence of a creaking noise can be verified in the laboratory by determining movement profiles and contact forces between the lock and the lock holder and the dirt and dust accumulation in the vehicle. This makes possible design improvements already at a very early stage in the development phase. Figure 8 depicts two design states of a lock. It was possible to eliminate the creaking by introducing furrowing in the contact area of the lock holder with the lock.

Figure 7: Control-force simulation for inner control with optimization loop
1 First optimization loop,
2 Second optimization loop.

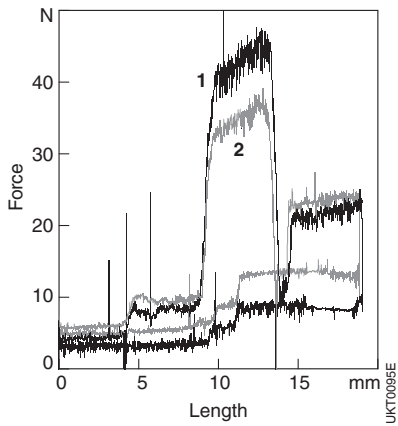
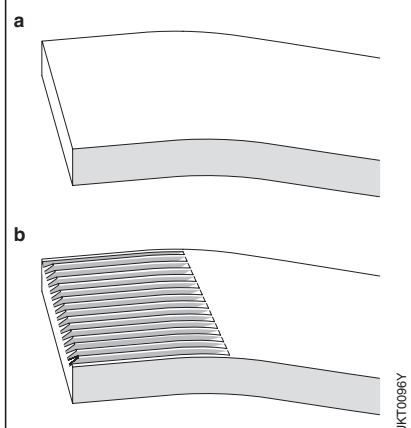


Figure 8: Design measure on the lock holder to reduce the creaking noise
a) Series state with creaking,
b) Optimized series state with furrowing (creak-free).



Safety functions

Behavior in a crash
The locking system must keep the doors securely closed during vehicle operation. In a crash situation high accelerations, forces, and deformations occur. These must not lead to the failure of the locking-mechanism/lock-holder connection and the unintended opening of the door. After the collision and in a rescue situation, however, the door should be easy to open. To keep the door closed during a crash, in many cases today mass locking mechanisms are installed in the outer door handles. These consist of a weight that is moved during customary accelerations in the crash and then block the outer door handle. This stops the handle from moving outwards and opening the door.

There are additionally tactilely operating systems. In these systems a lever attached to the lock touches the inside of the door outer panel. When this deforms, it blocks the outer control chain, and the door can then permanently not be opened from the outside. This too prevents unintended opening.

Improved systems have speed-based crash freewheels in the side-door lock. If the outer handle is operated too quickly, the movement comes to nothing and the door remains closed. This also covers crash situations in which acceleration-based systems such as mass locking mechanisms do not work. The door can be opened again after the crash.

Theft deterrence
The complete locking system must be protected against electrical, electronic or mechanical tampering. Mechanical and electrical connecting elements which are used to open the doors must not be accessible from the outside or may only be accessible with difficulty from the outside. Certain locks use magnetically operating Hall-effect sensors. These must be fitted in such a way that magnetic fields from the outside cannot cause unlocking or opening.

Many locks have a theft-deterrence system – also called double lock, dead lock or safe function. Depending on the manufacturer, this system is activated always or for instance by pressing the locking button on the remote control twice. Ordinarily this is not possible when there are persons in the vehicle. The theft-deterrence system prevents manual unlocking and opening by the inside door handle, for example after the side window has been smashed. This function is electrically activated in the lock by an additional motor. In the event of a power failure, the theft-deterrence system can be manually deactivated at least on one door via the lock barrel.

Detection of an opened door
The side-door lock also takes on the important task of signaling an open or closed door to the vehicle. This is done by microswitches in the lock which detect the position of the spagnolet and whether the locking pawl has locked. These signals are evaluated either in the door ECUs or in the car's central electronics. Alternatively, some manufacturers also provide door contact switches outside the lock, but these are less accurate.

This information is used to indicate an opened door or trunk lid in the instrument cluster. These signals are also used for the alarm system, automatic relocking of the doors after a certain period of time in the event of incorrect unlocking, or switching off the cruise control when the door is opened.

Development history

Historical view

The first locks to be used in vehicle doors were based on house locks. These so-called drop-latch locks could not compensate any tolerances, were able to absorb forces only in one direction, and had no preliminary detent. All the components were made of steel, the lever apparatus was not encapsulated, and there were no acoustic damping measures.

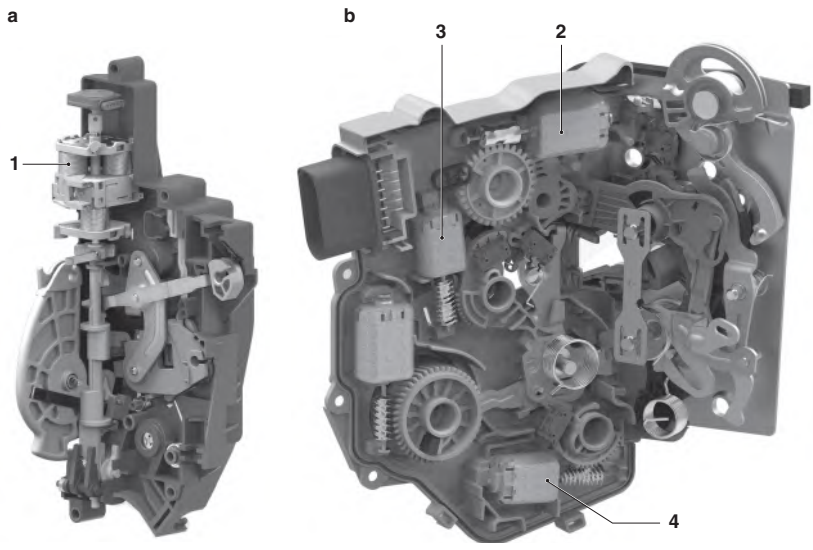
From around 1945 onwards this design principle was further developed into the so-called pin lock. This could absorb forces in several directions and had preliminary and primary detents. The locking actions was still performed directly by the drop latch.

In 1955 the first indirectly locking systems were introduced; above all, these reduced the control forces and at the same time improved tensile strength. The first acoustic damping materials reduced rattling noises. From 1960 the indirectly locking systems continued to make inroads into the market. For the most part the spagnolet was on the outside and the lock holder had a more complex design to ensure immobilization with buffer elements. Exclusively mechanical locks were used until the 1970s.

The first locks which locked and unlocked automatically were introduced in the USA in the 1960s. Electromagnets were used to activate them. Pneumatic or electrical control elements (actuators), which were screwed as separate units onto the mechanical basic locks, came later.

The increasing market share of central-locking systems led from the mid-1980s to the actuator being merged with the basic lock, and the so-called system lock was born. Plastic components and sensors were increasingly being used in its design. The indirectly operating locking mechanism with an internal spagnolet became the standard in the ensuing years.

Figure 9: Comparison of a conventional lock with a flex-pole lock
a) Flex-pole lock,
b) Conventional lock.
1 Flex-pole actuator for central locking, theft deterrence and electric child lock,
2 Motor for central locking,
3 Motor for electric theft deterrence,
4 Motor for electric child lock.



The first lock modules which brought together the outer and inner handles and the lock into a preassembled and pretested unit came onto the market in the mid-1990s.

Market trends

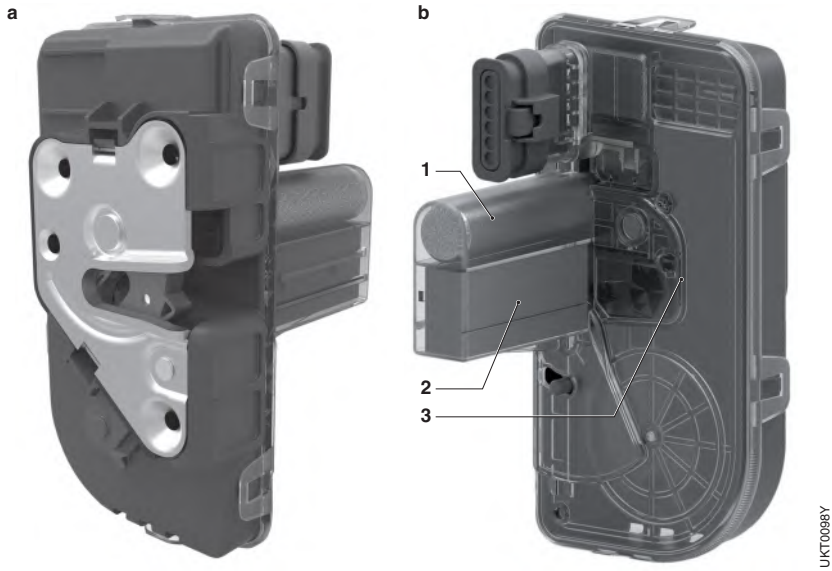
The mechanical complexity of locking systems increases with rising functionality. This results in increased weight and a variety of versions, both of which should actually be avoided. Possible solutions are reflected in two market trends: Simplifying the lock body by using a flex-pole actuator and simplifying the locking system by using an electric lock.

Today, a separate DC motor is used as a rule for each electromechanical function in a lock. This means – for a lock with central locking, theft deterrence, and electric child lock – the use of three motors. A flex-pole actuator is an electromechanical control element which can perform these three tasks. This eliminates the need for the motors together with the gear units and the lock becomes lighter, more compact, and more robust. Figure 9 shows a conventional lock and a flex-pole lock in comparison. The interfaces tot he handles and control points remain unchanged.

A further simplification is achieved if the conventional interfaces are also dispensed with, for example in the case of an electric lock. Here the locking pawl is lifted out only by an actuator, manual opening is no longer necessary. Sensors, switches or even smart phones can be used as controls. The lock body is also simplified such that the manual interface is dispensed with. Functions like central locking, theft deterrence, and electric child lock are now only effected by the control software.

Figure 10 shows an electric lock with a cable actuator for lifting out the locking pawl. With an electric lock it must be ensured that after a vehicle-battery failure or a power-supply failure it is still possible to open the door. The can be ensured by integrating a redundant voltage source (battery) into the lock.

Figure 10: Electric lock with cable actuator and electrical redundancy
a) Front view,
b) Rear view.
1 Integrated emergency power supply,
2 Integrated control,
3 Cable actuator.



Vehicle access by means of digital key

Task and design

In addition to the previously described components of a vehicle-access system these are extended for a system with a digital key. These additional components in the vehicle and the associated technology in the cloud and on the smartphone will be explained in the following.

A secure and digital vehicle-access system extends the described functions of a keyless-entry system by enabling drivers to unlock and lock their vehicle with their smartphone, entirely without the ignition key. The benefit of a digital key becomes apparent above all in fleets, which is why existing commercial-vehicle fleets, for example consisting of trucks, can also be equipment with the key-management system. This increases efficiency and the time available for the dispatchers and also for the drivers.

Technology in the vehicle

The hardware of this system ("Perfectly Keyless") consists of different ECUs which combine the technologies "Ultra Wide Band" (UWB), "Bluetooth Low Energy" (BLE), and "Near Field Communication" (NFC). What is new here is primarily the use of UWB technology, a radio-based communication technology for short-range use, which with its frequency band of 6.0 to 8.5 GHz facilitates a pure location accuracy of up to 20 cm. This also permits the implementation of so-called "Passive Entry", i.e., automatic unlocking and locking.

When a smartphone with the right key moves into the vehicle's communication range (communication zone), the smartphone is connected to it via BLE (Figure 11). As soon as the smartphone enters the unlocking zone, the vehicle automatically starts ranging via UWB and determines via the "time-of-flight" measurement (ToF) the distance between the smartphone and the vehicle. When the smartphone is located close enough to the vehicle, the vehicle is unlocked. The locking process works along similar lines.

In the vehicle the information received via UWB, BLE or NFC is managed and evaluated in the ECU which establishes vehicle access. This ECU is also the interface to the vehicle architecture. The received location and key information is made available by the corresponding antennas (Figure 12) via CAN link. Because the antennas are installed in different positions on the vehicle, the current position of the smartphone is determined via the UWB-based ToF measurement and the accompanying triangulation. In addition, the system is supplemented by an NFC

Figure 11: Automatic unlocking
UWB Ultra Wide Band,
BLE Bluetooth Low Energy,
NFC Near Field Communication.

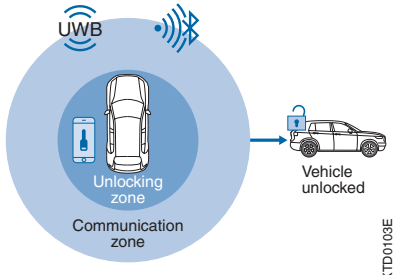
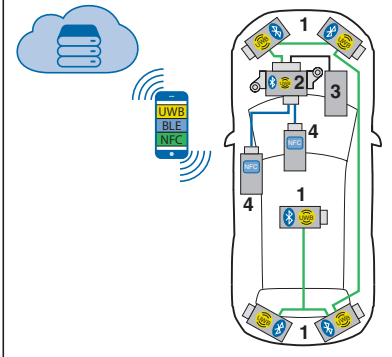


Figure 12: Antenna position in the vehicle
1 BLE and UWB antennas,
2 Master,
3 Body computer,
4 NFC antennas.



reader (Near Field Communication). This functions as a fallback level (back-up) and can even be used for vehicle access when the smartphone battery is flat.

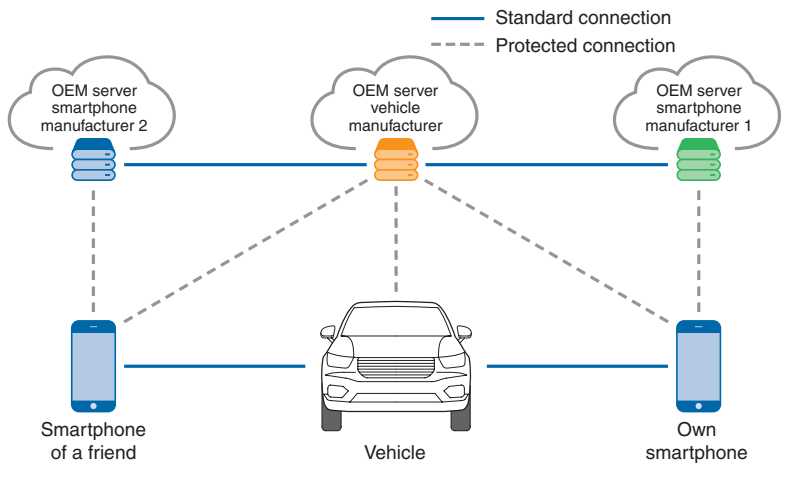
Technology in cloud and app

The digital "Key Sharing Service" is inspired by the Car Connectivity Consortium. The Car Connectivity Consortium (CCC) is a cross-industry organization that defines globally the technologies for smartphone-to-vehicle communication. The CCC is developing the digital key, a new standard which enables electronic devices (smart devices, e.g., smartphones) to interact with the vehicle (see Car Connectivity Consortium Administration, 2021).

The system architecture (Figure 13) shows the main components which are involved in creating a new "friend key", i.e., the key which is created by the owner of the vehicle and sent to an authorized driver. An owner key or a received friend key, i.e., a key already "shared" by the owner of the vehicle, is stored on the respective smartphone in the secure element. The smartphone uses the key app natively implemented by the smartphone OEM to communicate with the vehicle via NFC, BLE or UWB in order for example to unlock or lock the vehicle.

To make the benefit of the digital key available to fleets as well, the system also contains the option of a front end, with which a dispatcher for example can send keys to their drivers. The dispatcher can determine the period of time for which the digital key is valid and also withdraw keys that have already been issued. The back end manages the data for vehicles, drivers, smartphones, and ECUs in the vehicle. It is the interface to the front end. The back end is where the digital keys are created and distributed to the smartphones of the drivers. The digital keys are created in accordance with the CCC (Car Connectivity Consortium). The back-end components run in the Bosch IoT cloud (Internet of Things).

Figure 13: System architecture of digital keys (Car Connectivity Consortium, 2020)
OEM Original Equipment Manufacturer.



References
[1] FMVSS 206: U.S. Department of Transportation National Highway Traffic Safety Administration, Laboratory Test Procedure for FMVSS No. 206.
Door Locks and Door Retention Components (TP-206-08 February 19, 2010).
[2] ECE R11: Uniform provisions concerning the approval of vehicles with regard to door latches and door retention components. Addendum 10: Regulation No. 11 Revision 3.
[3] CISPR 25: Vehicles, boats and internal combustion engines – Radio disturbance characteristics – Limits and methods of measurement for the protection of onboard receivers. Edition 4.0, 10/2016.
[4] Dr. U. Nass: Erfolgsfaktoren global einsetzbarer Seitentürschlösser. ATZ Automobiltechnische Zeitschrift Issue 7–8/2008, Springer Automotive Media.
[5] Dr. U. Nass: Innovativer Zugang für zukünftige Fahrzeuggenerationen. Conference Transcript 16th Car Symposium 11 February 2016.
[6] Dr. U. Nass, J. Schulz: Conference Transcript VDI Conference Türen und Klappen (Doors, Gates, and Lids) 21–22 April 2015.
[7] Dr. U. Nass, J. Schulz: Conference Transcript VDI Conference Türen und Klappen (Doors, Gates, and Lids) 7–8 April 2017.

Anti-theft systems

Once the driver has left the vehicle, the vehicle must be protected against unauthorized use. It must be locked and the windows must be closed. This is expressly determined in the German road traffic regulations (StVO) in §14 s. 2 [1].

According to §38a of the German road traffic licensing regulations (StVZO, [2]) motor vehicles must furthermore be equipped with a security device to protect against unauthorized use. Vehicles are for this purpose equipped with a steering lock (steering-wheel lock), which is for the most part connected to the ignition lock. When the ignition key is removed a bolt is enabled which then engages in a hole in the steering column when the steering wheel is moved. This renders the vehicle unsteerable. When the ignition key is inserted again the bolt is pulled out when the steering wheel is turned and the steering is enabled again.

This does not prevent the engine from being started by short-circuiting the ignition. Protection against this is provided by the electronic vehicle immobilizer.

A car alarm is designed to make stealing a vehicle or breaking into the vehicle difficult by issuing an acoustic and optical alarm.

Electronic vehicle immobilizer

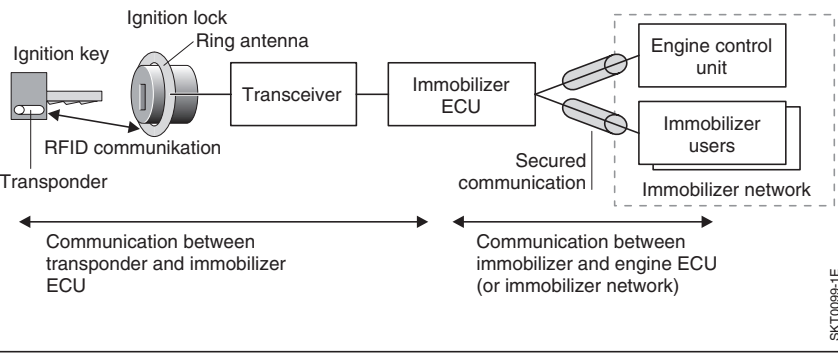
The electronic vehicle immobilizer is a theft-deterrence device which is designed to prevent the vehicle from moving under its own power if the associated ignition key is not inserted. Without authentication (checking of an asserted identity) of the transponder integrated in the ignition key, the immobilizer prevents operation of the vehicle by preventing for example engine starting, gear selection, or release of the steering lock.

Since 1998, in accordance with EU Directive 95/56/EU [3], all passenger cars sold in EU countries have been required to be fitted with an electronic immobilizer. Similar regulations apply in many other countries. The electronic immobilizer is not to be confused with automotive alarm systems or with remote-keyless entry systems for remote control of the door locking system.

System design

The electronic immobilizer (Figure 1) is integrated in the vehicle electronics and consists of the following components.

Figure 1: Immobilizer system



Transponder
The transponder is integrated in the ignition key and consists of an antenna, a microprocessor for signal processing and data transfer, a non-volatile memory (EEPROM) for storing a cryptographic key, and a capacitor as an energy accumulator. Passive RFID transponders (Radio-Frequency Identification) in the long-wave range (125 kHz or 134.2 kHz) are typically used in the automotive industry. These do not have their own energy source and their range is limited to a few centimeters [4].

Transceiver
The transceiver (also called the "reader") is connected to a ring antenna which on account of the short range of the radio signal is located near to or directly on the ignition lock. The transmit/receive unit of the transceiver transmits and receives the radio signals via the antenna and relays them to the immobilizer ECU.

Immobilizer ECU
The immobilizer ECU (also called the immobilizer server) contains a microcontroller for performing the immobilizer-function and communicates both with the transceiver and with the engine ECU, and if necessary with further ECUs in the immobilizer network. The immobilizer function can, instead of being integrated in its own immobilizer ECU, also be integrated in another ECU, e.g., in the instrument cluster, in the on-board computer, or in the engine ECU.

The immobilizer function has the task of checking the authenticity of the ignition key and the transponder and enabling the immobilizer.

Engine ECU
The engine ECU receives the signal to enable the immobilizer from the immobilizer ECU via a cryptographically secured connection. Different switch-off paths are disabled without an active immobilizer enable signal, meaning that engine starting is not possible. Typical switch-off paths are starter, fuel pump, and fuel injection.

Immobilizer network
To enhance security the ECUs of further vehicle components are incorporated into the immobilizer network, e.g., the automatic transmission (gear selection or opening of the parking lock can thus be prevented) or the steering.

Users of the immobilizer network must satisfy heightened security requirements, including securing bus communication and measures to increase protection against manipulation ("hardware security").

Functional description
Communication between transponder and immobilizer ECU
Communication between immobilizer ECU with transceiver and transponder is effected via a radio link and is activated as soon as the ignition key is in range or when the ignition is turned on. The transceiver generates with its antenna an electromagnetic field, which is received by the transponder antenna. Through inductive coupling the transponder receives it energy, which is stored in a capacitor.

Data are transferred as a rule in a half-duplex process: Data from the transceiver to the transponder (downlink) and data from the transponder to the transceiver (uplink) are transferred in succession. The received data are processed by the transponder's IC (integrated circuit). To return data to the transceiver, the transponder does not generate its own electromagnetic field, but instead modulates the transceiver signal by means of load modulation. The transceiver registers the fluctuations and decodes them as a transponder data signal [4].

The aim of this communication is to authenticate the transponder in the ignition key. Here the authentication can be unilateral (only the transponder authenticates itself to the immobilizer) or reciprocal (both the transponder and the immobilizer authenticate themselves to each other).

Authentication is performed in modern immobilizer generations according to the challenge-response scheme (see Figure 2). The prerequisite for this scheme is that the transponder and the immobilizer have a common, secret key K. The immobilizer transmits a challenge to the

transponder, which in turn converts and returns this by means of a cryptographic ENCRYPT function and the secret key K. Then the immobilizer performs the same calculation and compares the results (VERIFY). If the results match, the transponder has successfully authenticated itself to the immobilizer, i.e. proven it authenticity. A random number is typically used as the challenge with the aim of preventing attacks by recording and repeating, so-called replay attacks (see section entitled “Security”).

Communication between immobilizer and immobilizer users

The enable signal is transmitted by way of a cryptographically secured communication via the vehicle bus system to the engine ECU and if necessary the other users of the immobilizer network. Once the transponder in the ignition key has been successfully authenticated, the immobilizer is enabled by the immobilizer ECU and normal vehicle operation is possible.

The immobilizer is automatically activated again after the ignition key is removed.

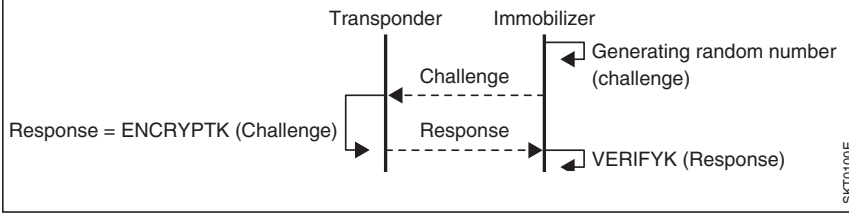
Security considerations

Both older and current generations of immobilizer systems have been successfully attacked in different ways, thereby showing up their weaknesses. Common weaknesses are incorrect implementations, deviations from standardized, cryptographic algorithms and processes, and the use of key lengths that are too short [5], [6], [7].

One requirement that is imposed on an immobilizer system is that the immobilizer even with great criminal application cannot be surmounted or can only be surmounted with great technical effort and financial expense. Attacks should furthermore not be scalable, i.e., knowledge of a successful attack on an individual vehicle should not be reusable for other vehicles. This is achieved among other things by the use of vehicle-individual or component-individual cryptographic keys.

To protect against so-called replay attacks, i.e., eavesdropping on and storing the communication for later use, a new random number (challenge) is generated in the challenge-response scheme for each authentication attempt. A valid authentication message that has been

Figure 2: Challenge-response scheme for transponder authentication
$ENCRYPT_K$: Encryption with the secret key K.
$VERIFY_K$: Comparison with internal calculation with the same key K.
Immobilizer is enabled if authentication successful.



eavesdropped on once is thereby rendered invalid for a later authentication attempt and can no longer be used.

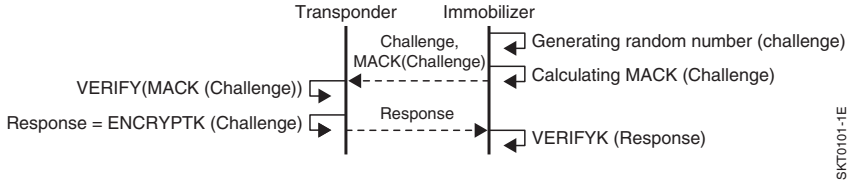
Once limitation of the challenge-response protocol shown in Figure 2 is that the transponder cannot check the authenticity of the immobilizer. Figure 3 shows an extension of the existing protocol. The immobilizer generates a random number (challenge) and then calculates its MAC (Message Authentication Code), a checksum for authenticating the challenge. Both the challenge and its MAC are communicated to the transponder. The transponder checks the MAC and can thus ensure that the challenge comes from a known, authenticated immobilizer ECU. Only if the MAC is valid does the transponder proceed with the protocol as described above; otherwise the communication is aborted at this point. This prevents the transponder from reacting to any challenge and sending back the associated response [8].

Production and infrastructure requirements

The use of symmetrical, cryptographic processes imposes heightened requirements on the overall product life cycle of each component relevant to the immobilizer. In production the cryptographic keys must be generated and programmed in a secure environment. Components relevant to the immobilizer are reciprocally paired via their secret key so that when a part is exchanged the associated key must be programmed again and the new components "taught" again.

Further applications are the provision of replacement ignition keys (e.g., after loss) and the disabling of ignition keys after they are stolen. The actions can be authorized by the manufacturer in each case via a central key administration and a secured infrastructure and the secret keys made available.

Figure 3: Challenge-response protocol with reciprocal authentication
MAC_K (Message Authentication Code): Checksum for authentication.
VERIFY: Comparison with internally calculated checksum.
$ENCRYPT_K$: Encryption.
$VERIFY_K$: Comparison with internal calculation.
Immobilizer is enabled if authentication successful.



Theft-alarm system

Regulations
Theft-deterrent (alarm) systems must meet the requirements of ECE regulations R18 [9] and R116 [10]. An armed alarm system must be capable of issuing acoustic and visual alarm signals in the event of unauthorized intervention in the vehicle. Extended requirements result from insurance regulations in various countries.

Permissible alarm signals
– Intermittent sound signals
  (25 to 30 s duration, 1,800 to 3,550 Hz;
  min. 105 dB(A),
  max. 118 dB(A) at 2 m distance,
– visual flashing signals
  (max. 5 minutes).

System design
The alarm system comprises a central control unit, sensors that detect intervention, and an alarm siren (Figure 1).

Alarm detectors
– Door and hood contacts,
– interior monitoring,
– tilt sensor,
– self-monitoring of the alarm siren.

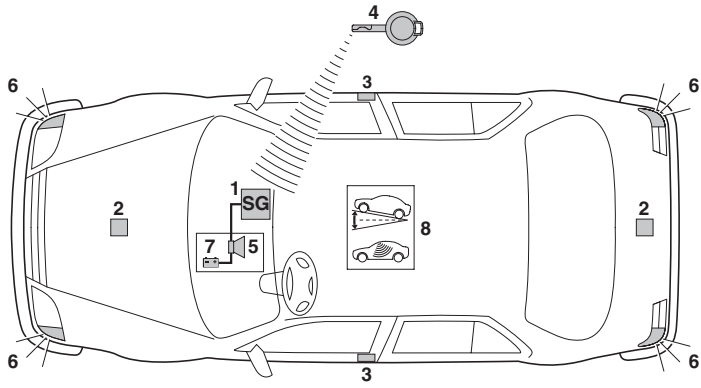
Alarm system control unit
Control of the alarm system is assumed by a control unit of the comfort electrical systems in the vehicle. The alarm system is armed or disarmed via the mechanical locks or via the radio remote control. When the command “Arm” is received, monitoring of the alarm detectors is activated. If required, selection buttons can be used to disable individual sensors. This deactivation applies once to the next “Arm” command.

Alongside the sensors, the alarm siren is also armed. Once the siren has assumed the “Armed” status, the control unit starts cyclical communication to monitor the lines. If an alarm is triggered, deactivation of the alarm (e.g. via the radio remote control) must lead to immediate termination of the alarm.

The control unit must ensure that the max. permitted number of alarms per alarm detector is not exceeded. Moreover, overlapping alarms must not lead to prolonged alarm output.

Alarm siren
The alarm siren comprises control-system electronics, a diaphragm as sound source, and a rechargeable battery for redundant power supply. The alarm tone is generated

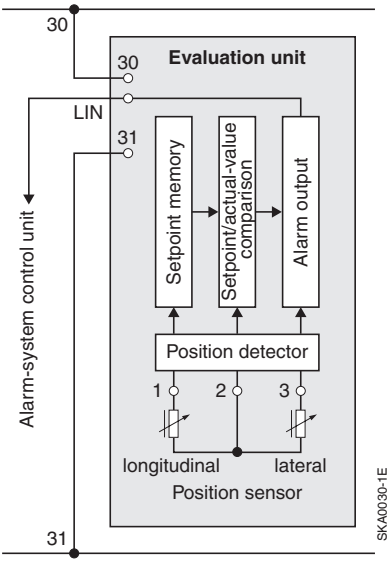
Figure 1: Theft-deterrent system
1 Alarm-system ECU, 2 Hood contact, 3 Door contact, 4 Key with radio remote control, 5 Siren, 6 Turn-signal lamps, 7 Battery, 8 Roof module with tilt sensor and interior monitoring.



by, for example, a piezo-loudspeaker. Communication with the control unit of the alarm system takes place across a serial single-wire bus (e.g. LIN). On the basis of a command from the control unit, the alarm siren can trigger the alarm or autonomously detect a manipulation. Here the following signals are monitored: cyclical communication, line contact to terminal 30 and terminal 31, voltage gradient of the vehicle battery, and overvoltage e.g. due to external supply. The aim is to recognize manipulations reliably and to prevent false alarms.

Tilt sensor
The task of the tilt sensor is to detect changes in the vehicle position caused by jacking up or towing the vehicle. A very sensitive acceleration sensor is deployed here. Its signal depends on the angle in relation to gravity. In order to monitor both the longitudinal and transverse axes of the vehicle, twin-axis micro-mechanical sensors are deployed (Figure 2). Particularly important is the reliable prevention of false alarms, as the vehicle rolling, e.g. due to wind, may not lead to an alarm.

Figure 2: Electronic wheel-theft and tow-away protection



Interior monitoring
The reflections of an ultrasound or a microwave field are analyzed. If the modification exceeds a comparative value, the alarm is triggered. Here, too, reliable prevention of false alarms is an important criterion.

References
[1] §14 StVO: Sorgfaltspflichten beim Ein- und Aussteigen (Duty of care when entering and exiting).
[2] §38a StVZO: Sicherungseinrichtungen gegen unbefugte Benutzung von Kraftfahrzeugen (Protective devices to prevent the unauthorized use of motor vehicles).
[3] Commission Directive 95/56/EC of 8 November 1995 adapting to technical progress Council Directive 74/61/EEC relating to devices to prevent the unauthorized use of motor vehicles (Text with EEA relevance).
[4] U. Kaiser: Digital Signature Transponder. In RFID Security: Techniques, Protocols and System-on-Chip Design, 2008.
[5] R. Verdult et al.: Gone in 360 seconds: Hijacking with Hitag2. In Security'12 Proceedings of the 21st USENIX conference on Security symposium, 2012.
[6] S. C. Bono et al.: Security Analysis of a Cryptographically-Enabled RFID Device. In Proceedings of the 14th conference on USENIX Security Symposium – Volume 14, 2005.
[7] R. Verdult et al.: Dismantling Megamos Crypto: Wirelessly Lockpicking a Vehicle Immobilizer. In 22nd USENIX Security Symposium (USENIX Security 2013). USENIX Association, 2013.
[8] S. Tillich et al.: Security Analysis of an Open Car Immobilizer Protocol Stack. In Lecture Notes in Computer Science Volume 7711, 2012.
[9] ECE-R18: Uniform provisions concerning the approval of motor vehicles with regard to their protection against unauthorized use.
[10] ECE-R116: Uniform technical regulations on the protection of motor vehicles against unauthorized use.

12 V vehicle electrical systems

Function and requirements

The vehicle electrical system of a motor vehicle with drive via an internal-combustion engine (classic drive) comprises the alternator as the energy converter, one or more batteries as the energy accumulators, and the electrical equipment as consumers. Figure 1 shows the schematic diagram of a vehicle electrical system. The nominal voltage of the battery is 12 V; this system is called a 12 V vehicle electrical system accordingly.

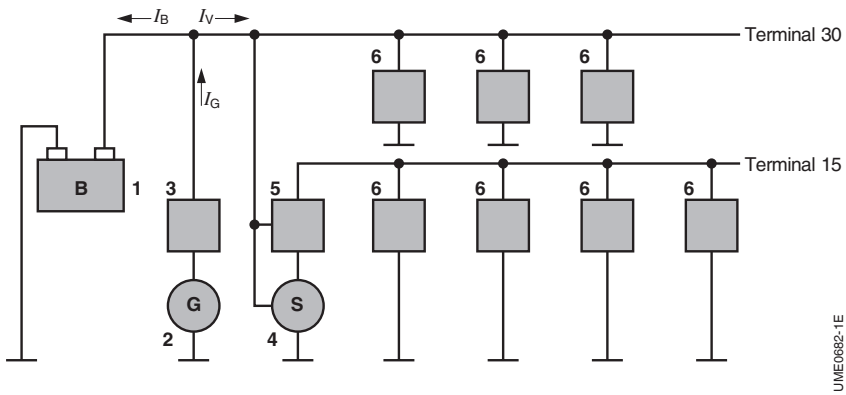
The energy from the battery is used to start the vehicle engine, hence the designation starter battery. This is performed via the starter in the classic drive systems. During vehicle operation, the ignition and fuel-injection system, the control units, the safety and comfort and convenience electronics, the lighting, and other equipment have to be supplied with power. An alternator supplies the required electrical power for these components and also to charge the battery.

Heightened comfort and safety requirements resulted and continue to result in a significant increase in the power demand in the vehicle electrical system. Added to this is the continuing trend towards the electrification of ever more vehicle components (e.g., seat adjustment, electric parking-brake system, electric steering servo). Driver-assistance systems and infotainment systems, which are still to be supplied with power even when the engine is stopped, have in the meantime found their way into the subcompact-car segment as well.

Electric and hybrid vehicles also have an electrical system for the electric drive as well as the 12 V vehicle electrical system. Here, however, the 12 V battery is supplied not by an alternator but instead by a DC/DC inverter, which transforms the high voltage from the additional electrical system into the voltage required for the 12 V electrical system (see Electrical systems for hybrid and electric vehicles).

Figure 1: Schematic diagram of a 12 V vehicle electrical system from the example of a one-battery electrical system

1 Battery, 2 Alternator, 3 Alternator regulator, 4 Starter,
5 Ignition switch, 6 Consumers.
I_B Battery current, I_G Alternator current, I_V Equipment current draw.
B Battery, G Alternator, S Starter.



Power of the consumers in the 12 V vehicle electrical system

Consumer classification

The electrical consumers in the 12 V vehicle electrical system have a variety of switch-on durations. Here the operative distinction is between:

– continuous loads, which are always switched on (e.g., electric fuel pump, engine management),
– long-term loads, which are switched on where necessary and then remain on for an extended time (e.g., low-beam headlamps, car radio, electric radiator fan),
– short-term loads, which are only switched on for brief periods (e.g., turn-signal lamps, stop lamps, electrical seat adjustment, power-window units).

Running-time-dependent electrical load requirements

The electrical load requirements encountered during vehicle operation are not constant. The first minutes after startup are generally characterized by high demand (e.g., heated rear window, seat heating, door-mirror heating), followed by a sharp drop in electrical load requirements. These consumers are switched off after a few minutes. The electrical load requirements are then determined predominantly by the continuous loads and the long-term loads. Table 1 shows examples of typical consumers.

No-load-current consumers

Different ECUs and consumers require a power supply even when the vehicle is parked. The no-load current is composed of the sum total of all these switched-on consumers. Most of these consumers switch off shortly after the engine is turned off (e.g., interior lights). Some, however, are always active (e.g., anti-theft alarm system).

An unexpectedly high no-load current can also occur when, for example, shutdown of the ECU interconnection does not work correctly or recurring “Wake-Ups” cause an ECU to frequently reactivate the interconnection when the vehicle is parked. The control unit networking must be configured and validated for this.

The no-load current must be supplied by the battery. The maximum value for the no-load current is defined by the vehicle manufacturers. Battery dimensioning is based among other things on this value. Typical values for the no-load current in a passenger car are approx. 3 to 30 mA.

Table 1: Installed consumers in a 12 V vehicle electrical system taking account of the switch-on duration (examples)

Electrical consumer	Power consumption	Average consumer power
Engine management with electric fuel pump	250 W	250 W
Radio	20 W	20 W
Side-marker lamps	8 W	7 W
Low beam (dipped beam)	110 W	90 W
License-plate lamp, tail lamps	30 W	25 W
Indicator lamps and instruments	22 W	20 W
Heated rear window	200 W	60 W
Fan for interior heating	120 W	50 W
Electrical radiator ventilator	120 W	30 W
Windshield wiper	50 W	10 W
Stop lamps	42 W	11 W
Turn-signal lamps	42 W	5 W
Fog lamps	110 W	20 W
Rear fog warning lamp	21 W	2 W

Design and operating principle

Schematic diagram

The electric system in a motor vehicle can be represented as the interaction of the energy converter (alternator), the energy accumulator (battery) and the consumers (Figure 1).

The alternator is driven via the V-belt by the crankshaft and converts mechanical energy into electrical energy. The alternator regulator limits the output power such that the desired voltage set in the regulator is not exceeded (14.0 to 14.5 V).

When the ignition key is removed, only a few consumers are supplied with voltage (e.g., anti-theft alarm system, car radio, auxiliary heating). The terminal via which these consumers are supplied is called “terminal 30” (continuous plus).

The other consumers are connected to “terminal 15”. When the ignition switch is in the “ignition on” position, the battery voltage is applied to this terminal so that now all the consumers are connected to the power supply.

Charge balance

The nominal output of alternators in the 12 V vehicle electrical system ranges from approx. 1 kW in subcompact-size cars to over 3 kW in the upper class. This is less than the consumers need in total. In other words, the battery must also supply power at times during vehicle operation. An equalized battery charge balance must always be guaranteed through the choice and dimensioning of battery, alternator, starter and the other consumers in the electrical system so that the internal-combustion engine can always be started and, when the engine has been switched off, specific electrical consumers can still be operated for an appropriate period of time.

The alternator supplies current (I_G, Figure 1) when the engine is running. To be able to charge the battery, the alternator must increase the vehicle-system voltage above the battery open-circuit voltage. However, the alternator is only able to do this when the cut-in consumers do not demand more current than it can supply. If the equipment current draw I_V in the vehicle electrical system is greater than the alternator current I_G (e.g., when the engine is idling), the battery is discharged. The vehicle system voltage falls to the voltage level of the battery from which current is drawn. If the equipment current draw I_v is less than the alternator current output I_G, a proportion of the current flows to the battery and acts as a battery charging current, I_B. The vehicle system voltage increases to the setpoint value specified by the voltage regulator.

The maximum alternator current is very much dependent on the rotational speed and the alternator temperature. When the engine is idling, the alternator can only supply 55 to 65 % of its nominal output. However, immediately after a cold start at low outside temperatures, the alternator is able from medium engine speed to supply the vehicle electrical system with up to 120 % of its nominal output. When the engine is hot, the engine compartment heats up, depending on the outside temperature and engine load, to 60 to 120 °C. High engine-compartment temperatures cause high winding resistances, which reduce the maximum alternator output.

Battery installation locations

In most cars the battery is located in the engine compartment. However, a large battery (e.g., 100 Ah) takes up a lot of space and sometimes, when the amount of space available in the engine compartment is restricted, may not be able to be installed there. Another argument against installing the battery in the engine compartment is the potentially high ambient temperatures. As an alternative, the battery may be installed in the luggage compartment.

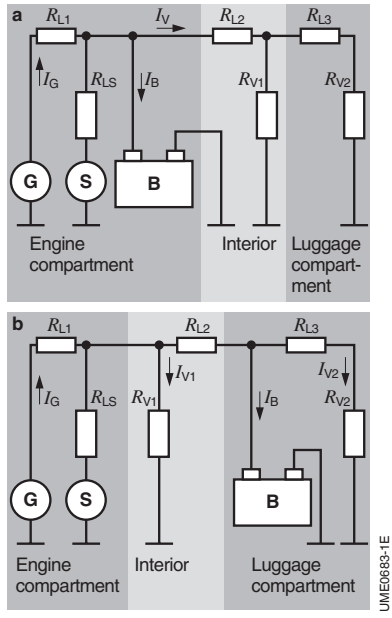
Effect of installation location on charge voltage

The line between a battery installed in the engine compartment and the alternator is shorter than for a battery installed in the luggage compartment. This affects the line resistance and thus directly the voltage drop on the cable. Voltage drops can be minimized by suitable conductor cross-sections and good connections with low contact resistance, even after a long service life.

Figure 2a shows the conditions for installation in the engine compartment. A battery installed in the luggage compartment requires a longer supply line with the additional line resistance R_{L2} (Figure 2b). Because of the higher voltage drop, the charge voltage for a battery installed in the luggage compartment is lower. The additional voltage difference caused by R_{L2} can be balanced by increasing the set-point alternator-voltage value. The power to be provided by the alternator increases as a result.

Effect of installation location on startability
Engine startability is dependent on the voltage applied at the starter. The higher this voltage, the higher the starter speed during the starting sequence. Because of the high starter current, the resistance of the supply line has a crucial effect on this voltage. In the case of the variant with the battery installed in the luggage compartment, the line between the battery and the starter is longer than for the battery installed in the engine compartment; accordingly, the resistance and with it the voltage drop are higher. Better startability is assured when the battery is installed in the engine compartment with short lines to the starter.

Figure 2: Battery installation locations
a) Installation in engine compartment,
b) Installation in luggage compartment.
G Alternator, B Battery, S Starter,
R_L Line resistances,
R_V Equipment resistances,
I_G Alternator current,
I_V Equipment current draw,
I_B Battery charging current.



Effect of ambient temperature
High temperatures, as can be encountered in the engine compartment, intensify aging effects in the battery (e.g., corrosion of the battery and water loss due to gassing), which have a negative impact on battery service life. High temperatures in the battery can be reduced by shielding.

The charge acceptance of the batteries is restricted at low battery temperatures, i.e., the battery takes in less charge, which can lead to a low state of charge. Low states of charge in turn facilitate aging effects such as sulfating, which lead to reduced usable capacity and therefore shorter service life.

Effect of installation location on voltage stability
Because the battery can only be charged with direct current, the alternating current generated in the alternator must be rectified. This job is performed by a diode rectifier integrated in the alternator (Figure 3). Rectification of the alternating voltage creates a pulsating direct voltage. Furthermore, the switching of the diodes – when the current is commutated from one diode to the next – generates high-frequency voltage oscillations, which are smoothed to the greatest possible extent by the interference-suppression capacitor installed in the alternator.

Electronic consumers (e.g., ECUs) can be disrupted or even damaged by the voltage peaks or the voltage ripple. The battery can use its large capacity to smooth the voltage fluctuations. However, because of the line resistance R_L between the alternator and the battery (Figure 3), they are not completely suppressed at the alternator. When the consumers are connected on

the battery side (Figure 4a) or after the battery (e.g., R_{V1} and R_{V2} in Figure 2a), they are supplied with the extensively smoothed system voltage. When the consumers are connected on the alternator side, i.e., directly to the alternator (Figure 4b), they are exposed to a greater voltage ripple and greater voltage peaks.

Electrical devices (consumers) which feature a high power input and relative insensitivity to overvoltage should therefore be connected close to the alternator, whereas voltage-sensitive loads with a low power input should be connected close to the battery.

Figure 4: Connection options for consumers

a) Battery-side connection of consumers,
b) Alternator-side connection of consumers.
G Alternator,
B Battery,
R_L Line resistance,
R_V Consumer resistance.

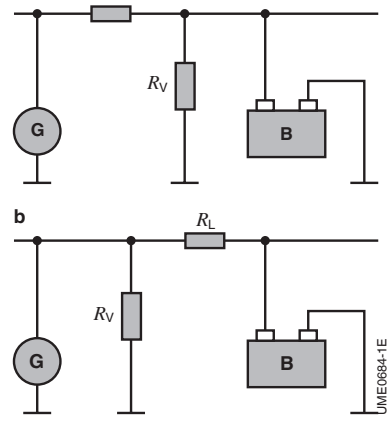
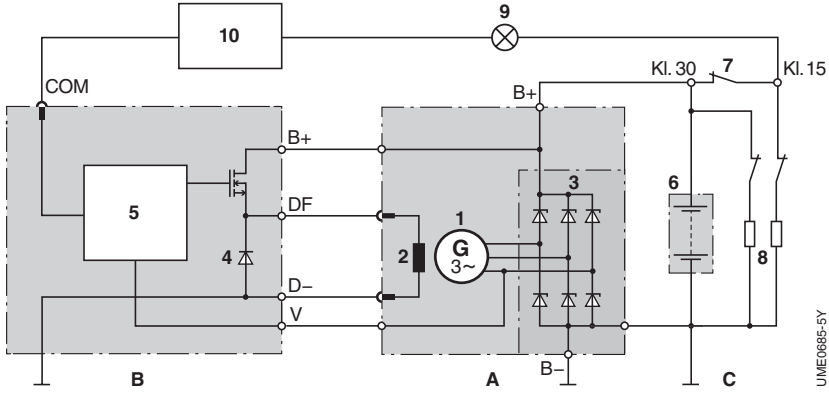


Figure 3: Interaction of alternator, alternator regulator and battery

A Alternator, B Alternator regulator, C Vehicle electrical system.
1 Stator with armature windings,
2 Rotor with excitation winding (power supply via carbon-brush/collector-ring system),
3 Rectifier diodes, 4 Free-wheeling diode, 5 Regulator logic, 6 Battery, 7 Driving switch,
8 Consumers, 9 Alternator indicator lamp, 10 ECU.
COM Communication interface (e.g., LIN bus),
DF Dynamo field, V Phase signal,
B+ Battery positive, B– Battery negative, D– Chassis ground.



Interaction of the components

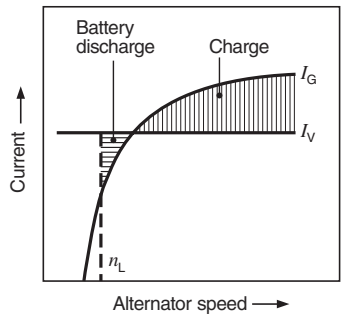
Current output of the alternator
Important component parts of the alternator are the stator (Figure 3) and the rotor, which is driven via the V-belt by the crankshaft. An alternating voltage is induced in the three stator windings (see Alternators) when a current flows in the rotor coil (excitation current) and a magnetic field is thereby built up. The excitation current is tapped from the generated alternator current (self-excitation). The induced voltage is dependent on the rotor speed and the level of the excitation current. The generated alternating voltage is rectified by diodes.

The voltage induced in the alternator is dependent on the alternator speed and thus on the engine speed. Therefore the voltage is low at low speeds. For an engine idling speed of n_L, the alternator can only supply some of its rated current if it has a conventional turns ratio (crankshaft speed to alternator speed) ranging from 1:2.5 to 1:3 (values for passenger cars; turns ratio is much higher for commercial vehicles) (Figure 5). The rated current is achieved at full load at an alternator speed of 6,000 rpm. The mean speed attained during vehicle operation must be sufficiently high in order to achieve the nominal alternator output. Driving cycles involving extensive levels of idling are particularly critical, because the available alternator output is so low that the battery is discharged when the electrical load requirements are high.

If the alternator voltage is higher than the battery voltage, a battery charging current flows into and charges the battery. The voltage is limited by the alternator regulator so as to maintain a system voltage of approx. 14 V.

Power generation by the alternator also affects fuel consumption. The increased consumption for 100 W output is in the order of 0.17 l per 100 km driven and is dependent on the efficiency of both the alternator and the internal-combustion engine.

Figure 5: Alternator current output I_G as a function of alternator speed
I_V Equipment current draw,
I_G Alternator current,
n_L Engine idle speed.



Voltage regulation in the vehicle electrical system
Generating the excitation field when starting
A magnetic field is required in the rotor to enable a voltage to be induced in the stator windings of the alternator. Self-excitation is not possible at low speeds after starting. In this phase the battery delivers the excitation current (external excitation).

The torque of an alternator running under load would hinder the starting sequence and idle stabilization of the internal-combustion engine. Alternator regulators therefore correct the excitation current during the starting phase to a low level (controlled pre-excitation). Current generation is delayed until the engine has run up to speed (Load-Response Start, LRS). Until this point the consumers are supplied by the battery.

Voltage regulation during vehicle operation
The excitation current is tapped from the generated alternator current (self-excitation). The alternator regulator adjusts the excitation field by way of a pulse-width-modulated (PWM) current in the rotor winding in such a way that the voltage at B+ (Figure 3) corresponds to the specified setpoint value. The frequency of the PWM signal is 40 to 200 Hz; the pulse duty factor is dependent on how much power is demanded by the consumers. In the case

of a load variation the vehicle system voltage changes, whereupon the alternator regulator adjusts the excitation field by adapting the PWM signal in such a way that the voltage is compensated.

The connection of the excitation winding is known as DF (dynamo field). The alternator regulator outputs the PWM signal as a DFM (DF monitor) to notify other ECUs of the alternator capacity utilization.

The alternator regulator requires the battery-voltage value in order to regulate. It obtains this value via terminal B+. In the case of a long supply line and high currents on the line between alternator and battery, the voltage difference between the battery and the regulator can be high due to the high voltage drop on the line to the extent that power generation by the alternator is too low and the battery is possibly insufficiently charged. This problem can be remedied by the S-terminal, which supplies the regulator with battery voltage via a cable which is separately connected to the battery positive terminal.

The alternator-regulator bus connection (e.g., via LIN bus) allows the system to vary the setpoint value to which the voltage is to be regulated. This makes possible functions such as, for example, recuperation, with which the setpoint value is increased in overrun mode so that the battery can be charged more effectively. The Load-Response Driving (LRF) function effects ramp regulation of the alternator voltage back to the setpoint value during vehicle operation after the connection of a high load and the associated sudden drop in the alternator voltage. This function prevents the alternator from abruptly subjecting the engine to load.

Charge indicator lamp

The charge indicator lamp is activated by the alternator regulator. It comes on with "ignition on" and goes out when the alternator is supplying current. The regulator activates the charge indicator lamp as soon as it detects a fault (e.g., alternator failure due to V-belt break, open circuit or short circuit in the excitation circuit, open circuit in the charging cable between the alternator and the battery).

Charging the battery

Due to the chemical processes that take place in the battery, the ideal battery charge voltage has to be higher at low temperatures and lower at high temperatures. The gassing voltage curve shows the maximum permissible voltage at which the battery does not produce gas significantly. The alternator regulator limits the voltage if the alternator current I_G is greater than the sum of the consumer equipment current draw I_V and the temperature-dependent, maximum permissible battery charging current I_B.

Alternator regulators are normally mounted on the alternator. If there are significant deviations between the temperatures of the voltage regulator and the battery electrolyte, it is better to monitor the voltage-regulation temperature directly at the battery. This is possible in vehicles with a battery sensor. The temperature value is conveyed by means of a communication interface (e.g., LIN bus) (see Battery sensor).

The arrangement of the alternator, battery and consumers influences the voltage drop on the charging cable and thus the charge voltage. The total current $I_G = I_B + I_V$ flows through the charging cable if all electrical equipment is connected to the battery. The charging voltage is accordingly less due to the high voltage drop. If all consumers are connected on the alternator side, the voltage drop is lower and the charging voltage is higher. The alternator regulator can take into account the voltage drop by measuring the actual voltage value directly at the battery via the S-terminal.

Design of the vehicle electrical system

Dynamic system characteristic curve

The dynamic system characteristic curve maps the relationship between battery voltage and battery current during a driving cycle (Figure 6). The envelopes reflect the interrelationships between the battery, alternator, and consumers in conjunction with temperature, engine speed, and speed ratio from engine to alternator. A large area in the envelope means that, with this type of vehicle electrical system, significant voltage fluctuations are occurring in the selected driving cycle and that the battery undergoes more powerful cyclization, i.e., its state of charge experiences powerful temporal changes. The system characteristic curve is specific to every different combination and every set of operating conditions, and is therefore a dynamic curve. Measuring systems can be connected to the battery terminals to plot the dynamic system characteristic curve.

Charge-balance calculation

By means of a computer program, the battery charge level is calculated from the consumer load and the alternator output power at the end of a specified driving cycle. Rush-hour driving (low engine speeds) combined with winter operation (low charging-current input to the battery) is regarded as a normal passenger-car driving cycle. The battery must maintain a stable charge balance even under these very unfavorable conditions for the energy balance of the vehicle electrical system.

The charge-balance calculation is used as the basis for defining the design of the alternator and battery.

Driving profile

The driving profile as an input variable for the charge-balance calculation is represented by the cumulative-frequency curve of the engine speed (Figure 7). It describes how often a specific engine speed is reached or exceeded.

In urban driving situations in commuter traffic, a car's engine is running at idle speed for a large part of the time due to the frequency of stops at traffic lights and consequently high traffic density.

A city bus running on scheduled routes has a high idle percentage time due to interruptions in driving at bus stops. Another factor that has a negative effect on the battery's charge balance is electrical consumers that are operated when the engine is switched off. Long-distance buses generally spend only a small proportion of the time with the engine at idle speed, but on the other hand may have periods when consumers with a high power input are operated with the engine switched off.

Figure 6: Dynamic system characteristic curves (envelopes for city driving)

1 Characteristic curve with large alternator and small battery,
2 Characteristic curve with small alternator and large battery.

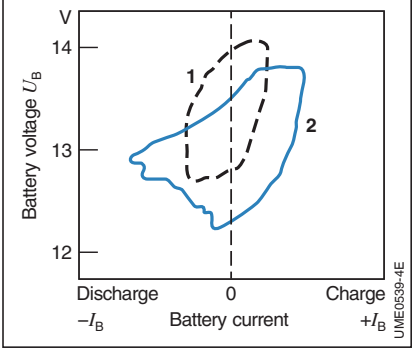
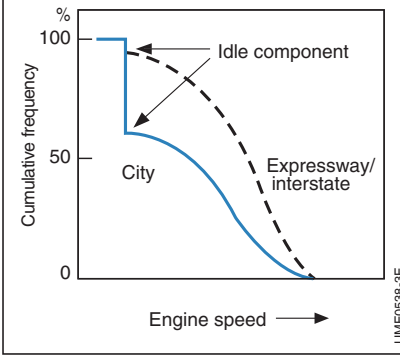


Figure 7: Cumulative frequency of engine speed in city and expressway/interstate driving



Vehicle-electrical-system simulation
In contrast with this non-analytical approach to charge-balance calculations, the situation in relation to the vehicle-electrical-system energy supply can be calculated at any time during operation using model-based simulations. They can also include electrical system management systems and assess their effectiveness.

In addition to adjusting current levels in the battery, it is possible to record the vehicle-system voltage characteristic curve and the battery cycle at any time during a trip. Calculations performed using vehicle-electrical-system simulations are always useful when it comes to comparing electrical-system topologies and assessing the effectiveness of consumers with a high dynamic response or only brief switch-on durations.

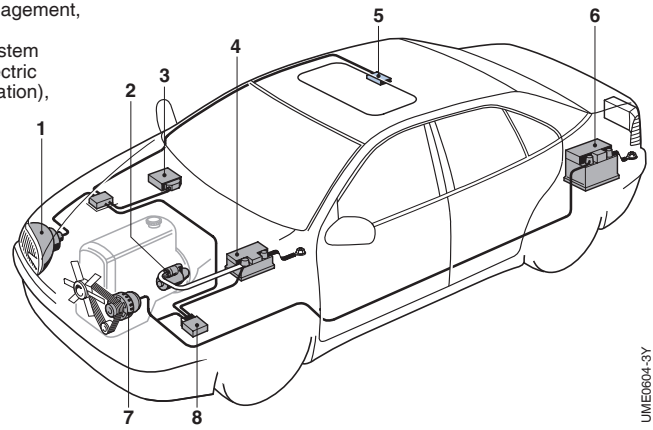
Electrical-system structures

One-battery vehicle electrical system
Figure 1 depicts a one-battery vehicle electrical system, as featured predominantly in passenger-car applications. Working as an energy accumulator is a battery which both supplies the current for the starting sequence and delivers the power supply to the consumers when the engine is stopped and thus there is no alternator output or insufficient alternator output in idling phases. This is the most widely used concept today in that it is the most cost-effective solution for supplying power in motor vehicles.

In the design of a battery for the one-battery electrical system which supplies both the starter and the other consumers in the vehicle electrical system, a compromise has to be found between different requirements. During the engine starting sequence, the battery is subjected to high current loads (300 to 500 A). The associated voltage drop has an adverse effect on certain electrical equipment (e.g., undervoltage reset on units with microcontrollers) and should be as low as possible.

By contrast, only comparatively low currents flow during standard vehicle operation. Battery capacity is the salient factor in providing a reliable current supply. Nei-

Figure 8: Two-battery vehicle electrical system
1 Lighting system,
2 Starter,
3 ECU for engine management,
4 Starter battery,
5 Further electrical-system consumers (e.g., electric sliding-sunroof operation),
6 General-purpose battery,
7 Alternator,
8 Electrical-system ECU.



ther properties – output nor capacity – can be optimized simultaneously.

Two-battery vehicle electrical system
In vehicle electrical systems with two batteries – starter battery and general-purpose battery – the "high power for starting" and "general-purpose electrical supply" functions are separated by the vehicle electrical system control unit (Figure 8) to make it possible to avoid the voltage drop during the starting process, while ensuring reliable cold starts, even when the charge level of the general-purpose battery is low.

Starter battery (starting battery)
The starter battery must supply a high amount of current for only a limited period of time (during starting). It is therefore designed for a high power density (high power for low weight). Compact dimensions allow installation in the immediate vicinity of the starter motor with short connecting cables (low voltage drop on the line). The capacity is reduced.

General-purpose battery
This battery only supplies the vehicle electrical system (excluding the starter). It supplies currents to the electrical-system consumers (e.g., approx. 20 A for the engine-management system) and must therefore be able to store and provide large amounts of energy. It also has a high cycle stability. This means it can be charged and discharged very often before the performance criteria of the battery are no longer met. Dimensioning is based essentially on the capacity reserves required for activated consumers, the consumers that operate with the engine switched off (no-load-current consumers, e.g., receiver for central-locking radio remote control, anti-theft alarm system), and the minimum permissible charge level.

Vehicle power supply control unit
The vehicle power supply control unit in a two-battery vehicle electrical system separates the starter battery and the starter from the rest of the vehicle electrical system provided this can be supplied with sufficient power by the general-purpose battery. It therefore prevents the voltage drop that occurs during starting, affecting the performance of the vehicle electrical system. When the vehicle is parked, this prevents the starter battery from becoming discharged by electrical equipment that draws current when switched on and by no-load-current consumers.

The 12 V battery can be charged at higher voltage than the usual alternator voltage in the vehicle. The voltage level is limited by consumers, e.g., the bulbs, whose service life severely deteriorates with the voltage. By separation of the starter battery from the remainder of the vehicle electrical system, there are theoretically no limits for the charge voltage level within the starter battery. It can therefore be raised to the "ideal" value by means of a DC/DC converter (e.g., 15 V). This is dependent on the temperature and on the state of charge. Charging duration is minimized with the increased charge voltage.

If there is no charge in the general-purpose battery, the vehicle power supply control unit is capable of provisionally connecting both vehicle electrical systems. This means that the vehicle electrical system can be sustained using the fully-charged starter battery. In another possible configuration, the vehicle power supply control unit for the starting operation would connect only the start-related consumers to whichever battery was fully charged.

Electrical-system parameters

State of charge
The state of charge of the battery (SOC) is one of the most important parameters in the vehicle electrical system. It can be defined as the ratio of the amount of charge still stored in the battery (actual state of charge) to the maximum amount of charge which the fully charged battery when new can store:

$$\text{SOC} = \frac{Q_{\text{act}}}{Q_{\text{max}}} .$$

The value Q_{max} is obtained when a fully charged battery is discharged at discharge current I_{20} – corresponding to a twentieth of the nominal capacity in amperes (5 A for a 100 Ah battery) – until the cutoff voltage of 10.5 V is reached. The amount of charge which was drawn during this discharge process corresponds to Q_{max}.

Since in this way Q_{max} is only accessible by means of one measurement (integral of the current over the time, i.e., current and time measurement), the definition can also frequently be provided by the nominal capacity of the battery, which can be found on the label, where: $Q_{\text{max}} = K_{20}$ (nominal).

The currently stored amount of charge Q_{act} is obtained from the difference between Q_{max} and the amount of charge drawn during the discharge of the fully charged battery.

The state of charge of the battery correlates directly with the electrolyte density, where furthermore the steady-state voltage of the battery is proportional to the electrolyte density. The voltage value which is obtained when a stable voltage final value is assumed after the battery charging or discharging process is referred to as steady-state voltage. This can, owing to the slow diffusion and polarization processes in the battery, take several days, especially after long charging phases. The steady-state voltage is measured at the terminals.

The state of charge is defined as follows:

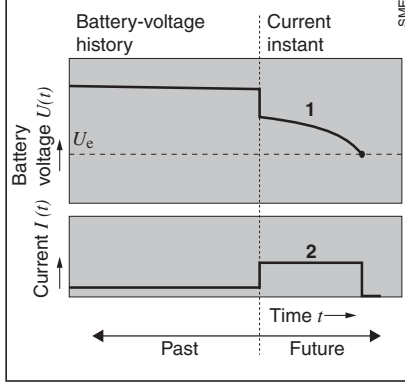
$$\text{SOC} = \frac{(U_{\text{current}} - U_{\text{min}})}{(U_{\text{max}} - U_{\text{min}})}$$

where
U_{current}: Current steady-state voltage.
U_{max}: Steady-state voltage of the fully charged battery (SOC = 100 %).
U_{min}: Steady-state voltage of the battery at SOC = 0 %. Because the dependence of the steady-state voltage on the state of charge is nonlinear for low states of charge (approx. less than 20 %), the value linearly extrapolated to SOC = 0 % must be used here.

In this way, it is possible to infer the state of charge from the measured steady-state voltage.

Functional and performance capability
The capability of the battery to meet a specified functional or performance requirement in its current state – e.g., providing the starting power for the internal-combustion engine – is called the SOF (state of function). The SOF is application-specific and can therefore not be defined generally. For example, the SOF can be used not only as a measure of startability, but also for evaluating the current performance capability of the battery to supply other electrical heavy current consumers, e.g., electrical power steering.

Figure 9: Determination of the SOF value by calculating the voltage curve under load
1 Predicated battery voltage $U(t)$ for the given virtual current profile $I(t)$,
2 Virtual current profile.
U_e Basis for the SOF calculation.



The current performance capability of the battery is evaluated using the SOF by precalculating the voltage drop at battery load with a predefined current profile. If the precalculated (predicted) battery voltage U_e drops below a specified threshold, this means that the battery is no longer delivering the required performance (e.g., for an engine start) (Figure 9).

Moreover, it is also possible to query the current remaining charge or the discharge duration which states how long the battery can continue to deliver the required performance. An application example of this is how long a vehicle can be parked at a known no-load current consumption without losing startability. In this case, the SOF does not provide the voltage drop at high-current load, rather the remaining charge reserve or the discharge duration until the startability threshold is no longer reached (Figure 10).

As already mentioned, the SOF is determined on the basis of the current battery state, i.e., it depends on the current state of charge (SOC), the battery temperature, and the state of health (SOH) of the battery. These variables have to be determined by the battery-status recognition (see Battery sensor) for SOF calculation.

State of health

Depending on the respective operating and ambient conditions, different aging processes take place in the lead battery used in the 12 V vehicle electrical system, which can, for example, lead to an increase in internal resistance (caused by corrosion) or to capacity losses (caused by active mass loss or sulfating) and therefore deteriorate the general performance and storage capability of the battery, i.e. its SOH (state of health) over the service life.

The state of health of the battery with regard to its storage capability can, for example, be expressed by the ratio of current capacity K_{20} to nominal capacity of the battery when new K_{20new}:

$$\text{SOH} = \frac{K_{20}}{K_{20new}} \cdot 100\ \%.$$

Further (application-dependent) SOH quantities can be derived from the corresponding SOF variables by calculating them not for the current battery state, rather with defined values for state of charge (e.g., SOC = 100 %) and battery temperature (e.g., 25 °C) to achieve the desired aim of only factoring the dependence on the state of health into the SOH quantity.

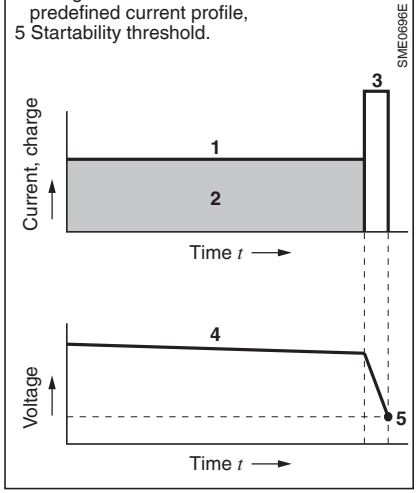
Whereas the SOF states how well a performance requirement can be met in the current state of the battery, the corresponding SOH variable states how well the battery will meet this performance requirement generally in the current state of health. For example, in this way an SOH variable which calculates the battery's remaining minimum performance capability for a start-stop function can be used to permanently deactivate this function when the battery is severely aged, in order to prevent the risk of breakdowns.

Alternator capacity utilization

The current flowing in the alternator excitation winding determines the voltage induced in the stator windings. The alter-

Figure 10: Calculation of the SOF value based on the available charge reserve

This calculation is used to determine the charge reserve or the time duration after which under no-load current the startability threshold will be reached.
1 No-load current,
2 Charge reserve,
3 Starting current,
4 Voltage curve under load with predefined current profile,
5 Startability threshold.



nator regulator regulates the required excitation current by means of a pulse duty factor (PWM signal). DF (dynamo field) is the terminal through which the excitation current is supplied (Figure 4). The pulse duty factor of the PWM signal indicates the alternator capacity utilization, i.e., whether the alternator still has reserves and can supply even more current for additionally connected loads.

The alternator regulator additionally outputs this signal as a DFM signal (DF monitor). Regulators with a bus interface apply this pulse duty factor to the bus (e.g., LIN bus). The excitation current is also output in amperes. Different ECUs evaluate the DFM signal in order, for example, to switch off the seat heating or the windshield heater when the alternator is subject to high capacity utilization.

Electrical energy management

Motivation

Reduced fuel consumption

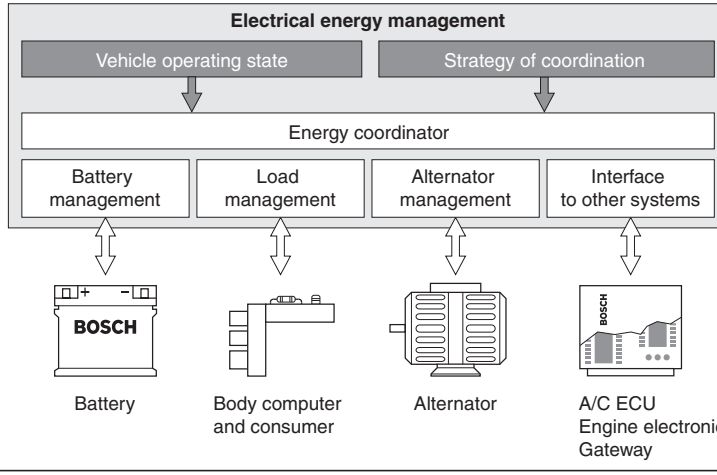
One of the primary goals of motor-vehicle manufacturers is to reduce fuel consumption and greenhouse gases, in particular CO_2. This is achieved by optimizing the energy flows in motor vehicles. Measures to achieve this goal include:

– Avoiding idling losses by using the start-stop function (automatic engine switch-off and restarting for example when waiting at traffic signals on red).
– Increasing the efficiency of electrical power generation by optimizing the alternator and applying intelligent alternator activation (recuperation).
– Electrically driven accessories to facilitate activation to suit demand by means of isolation from the internal-combustion engine.

Electrical power demand

Additional comfort and convenience functions and electrified accessories give rise to increasing electrical power requirements and at the same time to a reduced range of speed for electrical power generation (e.g., due to start-stop operation). New comfort, convenience and safety functions (e.g., electrical power steering, electric water pump, PTC auxiliary

Figure 11: Electrical energy management (EEM)



heater, electric climate control in vehicles with start-stop function) require additional electrical power, so much so that it is useful to integrate an electrical energy management system (EEM).

Purpose of electrical energy management

Electrical energy management controls the energy flows and at the same time safeguards the electrical power supply in order to maintain vehicle startability and to reduce breakdowns caused by flat (discharged) batteries. Electrical energy management also stabilizes the battery voltage and optimizes the availability of comfort and convenience systems – even when the engine is stopped. This can be achieved by safeguarding a positive or at least equalized charge balance during vehicle operation and by monitoring the power demand when the engine is stopped. Furthermore, peak loads can be reduced by the coordinated switching of electrical consumers. This is coordinated in the electrical energy management system (Figure 11).

The effects of the measures taken sometimes conflict with one another. For example, switching off comfort and convenience consumers results in loss of comfort and convenience, while disabling the start-stop function results in increased fuel consumption. Different vehicle manufacturers prefer different measures, and the possible measures for safeguarding the charge balance are prioritized accordingly.

Functions of electrical energy management

No-load-current management

The no-load-current management implemented in the battery sensor by means of software regularly monitors the battery status and thus startability when the engine is stopped. With the aid of precise battery-status recognition it is possible to optimize the availability of consumers by means of no-load-current management, i.e., the operating time of the comfort and convenience consumers can be maximized. In the event of an imminent loss of startability, electrical energy management can, for example, send a message to the display module in order to notify the user. In addition, when the startability limit is approached, load management will reduce the power consumption (e.g., by reducing the power consumption of the A/C fan) right down to shutting down individual consumers in order to maintain startability for as long as possible. Examples of such comfort and convenience consumers are the auxiliary heater, infotainment, navigation system, and radio.

Energy management during vehicle operation

The function of electrical energy management with the alternator active is, in addition to load management, principally alternator management including a recuperation function and the energy-management interface to other systems such as, for example, engine management.

Switching of consumers

Load management coordinates the activation and deactivation of consumers in order to reduce power peaks. Load management is also involved in controlling the high-performance heating systems (windshield heater and PTC auxiliary heater).

In vehicle operation, too, safeguarding restartability is the primary function of electrical energy management. In critical battery states, load management reduces the electrical power demand in order to recharge battery as quickly as possible. Reservoir-type comfort and convenience consumers (heating systems) are switched back by preference, because intelligent activation can be used to delay for as long as possible perceptible deviations from the nominal performance.

There are limits to deactivating comfort and convenience functions in that this will only be accepted by the user in rare exceptional circumstances. The vehicle electrical system must therefore be configured in such a way that these situations only occur rarely. Noticeable effects must be indicated to the user in order to explain the performance that is deviating from normal operation.

Increasing the alternator output

As an alternative or complement to reducing the electrical power demand, it is possible by increasing the engine speed to increase the electrical power generation by

the alternator (e.g., idle-speed increase or disabling of engine stop during start-stop operation). For example, to increase the idle speed, electrical energy management issues a request via the data bus to the engine management. These measures have a direct influence on fuel consumption and noise generation; they must therefore be optimally matched to the individual vehicle.

During a the vehicle's kinetic energy is converted at least partially into electrical energy and delivered to the battery for storage. This function requires an alternator that can be controlled via an interface to input the desired alternator voltage and a battery sensor for recognizing the battery status. The function of the battery sensor itself can be partitioned in the engine electronics, a gateway or a body computer.

Recuperation

During overrun fuel cutoff an increased desired voltage is input for the alternator in order to charge the battery intensively. Electrical power is generated at this operating point without fuel consumption. In driving states with poor efficiency in terms of electrical power generation the alternator voltage is decreased and the battery slowly discharged again in order to minimize the fuel demand for electrical power generation.

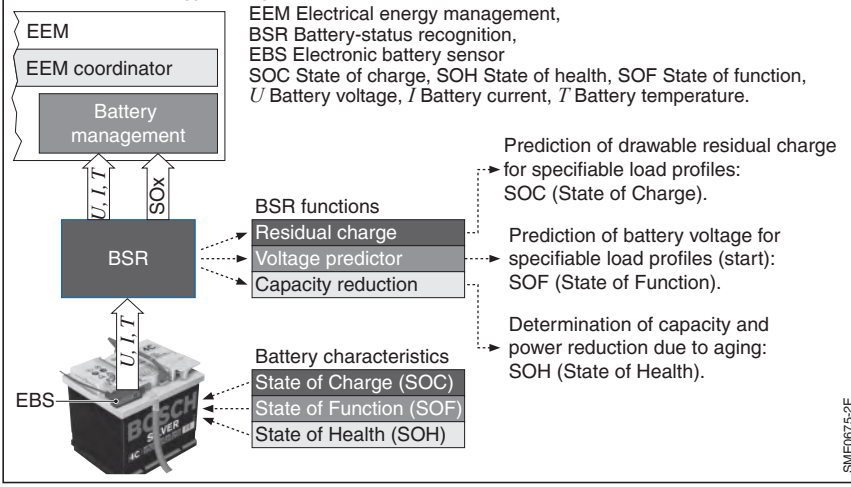
A fully charged battery cannot absorb charge. For this reason, recuperation is only possible with a partially charged battery (partial state of charge, PSOC). This is a departure from the conventional charging strategy, the objective of which is to achieve as fully charged a battery as possible. A minimum battery status required for startability must not under any circumstances be undershot, i.e., the current battery status must be known to electrical energy management.

The recuperation function results in an increased cyclical variation of the battery, the influence of which on battery aging must be tested on an application-specific basis. The use of AGM batteries (absorbent glass mat) to increase the possible energy throughput (throughput in Ah over the service life, the throughput critical to service life increases by a factor of 3) is therefore recommended.

The recuperation algorithm must take into account the influence of voltage changes on the consumers, since these can be observable (e.g., change in the speed of the A/C fan or light flickers).

Recuperation provides for fuel savings in the range of 1 to 4 %, depending on the cycle and configuration of the function.

Figure 12: Interaction of battery sensor, battery-status recognition, and electrical energy management



Battery-status recognition and battery management

Function

A crucial element to sound electrical energy management is a battery-status recognition facility (BSR) which reliably calculates the capability of the battery. Algorithms for battery-status recognition normally use as input variables the measured variables of battery current, voltage and temperature. On the basis of these variables the state of charge (SOC), the state of function (SOF) and the state of health (SOH) of the battery are determined and made available to electrical energy management as input variables (Figure 12).

A battery sensor which measures the battery current and voltage directly is used to measure the battery variables. The battery temperature is determined by way of a temperature measurement in the vicinity of the battery, since direct measurement of the battery electrolyte temperature in the vehicle would require an intervention in the battery, which is currently not possible.

Example

An example of a battery-status-recognition function is determining startability on the basis of the SOF. With the SOF the future behavior of the battery when subjected to starting current is predicted. In other words, battery-status recognition determines the battery-voltage dip for a given starting-current profile (Figure 9). Because the minimum voltage level for a successful start is known, the predicted voltage dip provides a measure of current startability. Depending on the interval between the predicted voltage dip and the startability limit, electrical energy management defines measures for maintaining or improving startability.

Battery sensor

The battery measured variables of current, voltage and temperature must be sensed very precisely, dynamically and synchronously. Particularly the measurement of currents in the range of a few mA through to starting currents of more than 1,000 A makes great demands on the sensor system. The electronic battery sensor (EBS) is mounted directly on and combined with the battery terminal. Because the terminal compartment is standardized in accordance with DIN 50342-2 [1], application on different batteries is not necessary.

Current is measured using a special manganin shunt. The core of the battery sensor's electrical wiring is an ASIC which contains among others a powerful microprocessor for recording and processing measured values. The battery-status-recognition algorithms are also processed on this microprocessor. Communication with higher-level ECUs is via the LIN bus.

As well as calculating the battery status for electrical energy management, the battery sensor can also be used for other functions. For example, precise sensing of current and voltage can also be used for prompted troubleshooting in production and workshops (e.g., searching for faulty no-load-current consumers).

References

[1] DIN EN 50342-2 (2008): Dimensions of batteries and marking of terminals.

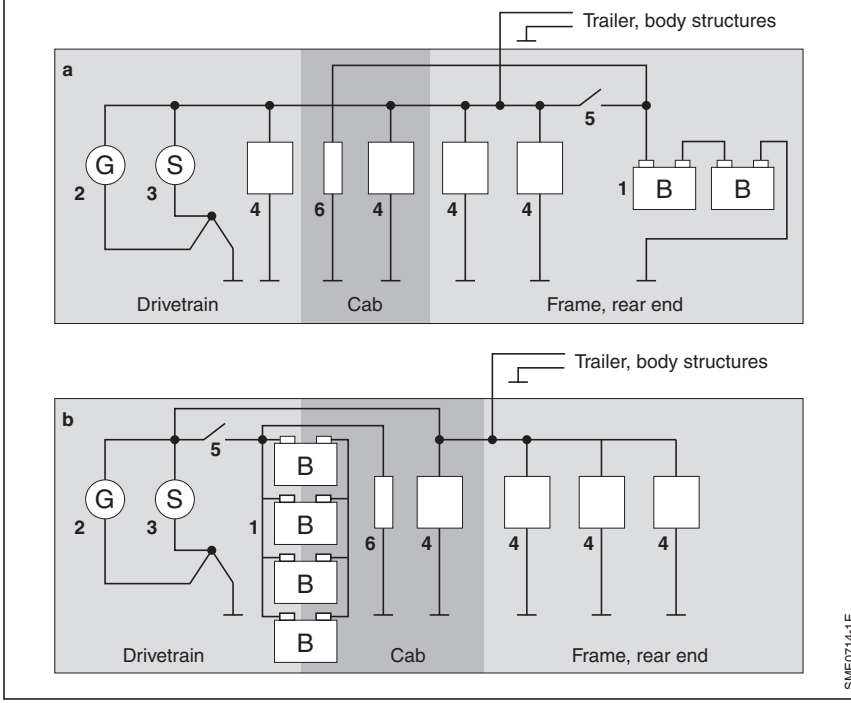
Vehicle electrical systems for commercial vehicles

Requirements of truck electrical systems

As in passenger cars, the majority of electrical consumers in trucks and buses are supplied with power via a low-voltage electrical system. It takes a much greater amount of power to start the large diesel engines of trucks (the majority of which are still used as the drive) than it does to start passenger-car engines. The nominal voltage of truck electrical systems in most regions around the world is therefore 24 V. 12 V vehicle electrical systems are sometimes used in light-duty vehicle classes (e.g., 3.5 to 7.5 t) and in heavy-duty vehicles in certain regions, e.g., North America.

Aside from the voltage, which is primarily defined by the vehicle's target market, a range of factors imposes further requirements on truck electrical systems:
– Vehicle-fleet operators: vehicle availability, “total cost of ownership” (TCO).
– Drivers: availability of comfort and convenience consumers.

Figure 1: Schematic diagram of truck electrical systems
a) 24 V electrical system with two series-connected 12 V batteries,
b) 12 V electrical system with several parallel-connected batteries.
1 Batteries, 2 Alternator, 3 Starter, 4 Electrical consumers at terminal 15,
5 Isolating switch (optional), 6 Continuously supplied consumers.



– Variety of vehicle types and use: number and type of consumers, power requirements, vehicle utilization, and driving profiles.
– Statutory requirements for example governing the carriage of dangerous goods (ADR, Accord européen relatif au transport international des marchandises dangereuses par route - European Agreement concerning the International Carriage of Dangerous Goods by Road), tachographs, etc.

The requirements imposed by vehicle operators and drivers differ between trucks and passenger cars in terms of the prioritization. For commercial vehicles, the availability of the vehicle and the costs come first. High requirements for example are very often imposed on the cold-starting performance or on the ability to drive the vehicle to a repair shop independently after initial faults (“limp home”). Further examples are competing requirements relating to fuel consumption, vehicle availability and battery life, which must be taken into account in energy-management functions for trucks. Requirements relating to the availability of comfort and convenience consumers only come second in the order of priorities; these requirements – in contrast to passenger-car requirements – often relate to the power supply when the vehicle is stationary, e.g., running an electrical independent (stationary) air conditioner while sleeping in the vehicle overnight.

The great variety of vehicles gives rise to requirements relating to the scalability and adaptability of the electrical systems to cope with different vehicle equipment, usage profiles, and body structures.

The average power requirement of basic vehicles is typically below that of passenger cars, e.g., for tractor units including semitrailer approx. 500 to 700 W. The reasons for this are the lower number of comfort and convenience consumers and the use of compressed air for many actuators. Body structures can however increase the power requirement.

Schematic design of truck electrical systems

24 V vehicle electrical systems

Similarly to passenger-car electrical systems, the electrical system of a truck consists of one or more energy converters (alternator(s) or in electric vehicles a DC/DC converter of HV (High Voltage) to 24 V), two series-connected 12 V batteries (typically 2 × 100 to 225 Ah capacity), and the electrical consumers (Figure 1a). The functions of the batteries are:
– to supply consumers while the vehicle is stationary (e.g., when sleeping in the vehicle overnight or to cover the no-load current demand),
– to provide power for engine starting in conventional vehicles, and
– to give short-term assistance to the alternator or the DC/DC converter during peak loads when driving.

Installation locations for batteries

Typical installation locations for the batteries are at the sides on the frame or inside the frame at the vehicle rear end (often the case on tractor units), which in conventional vehicles necessitates long cable lengths from battery to starter and alternator.

The positive cable is typically fashioned by way of a main power distributor. Conducting elements of the vehicle, e.g., engine block, clutch or transmission housing, and sometimes the vehicle frame, are used for the ground cable. However, particularly instead of the frame being used, sections are also routed via separate ground cables.

Isolating switches can additionally be fitted to isolate parts of the electrical system; these are sometimes required by law (e.g., for dangerous-goods vehicles).

12 V vehicle electrical systems

The design of 12 V electrical systems for trucks is essentially the same as the design of 24 V electrical systems. On light-duty vehicles (e.g., 3.5 to 7.5 t) a battery is mounted on the frame (e.g., a larger passenger-car battery (100 Ah) or a small truck battery (100 to 140 Ah). On heavy-duty vehicles a number of 12 V batteries are connected in parallel. On vehicles for

the North American market it is customary for example to install up to four batteries behind the engine and transmission under the cab and optionally up to four further batteries at the side on the frame. Figure 1b shows a typical schematic design of 12 V electrical systems for heavy-duty trucks.

24 V/12 VDual-voltage electrical systems

Insular electrical systems for 12 V / 24 V small consumers

Frequently 12 V components for passenger cars are used in European trucks and 24 V components from other regions are used in North American trucks. In such cases insular electrical systems are created via 24 V/12 V or 12 V/24 V DC/DC converters. These insular electrical systems are typically not supported by batteries, i.e., the DC/DC converters must be designed for the peak-power requirements and any return of current from components must be backed up by sufficient input capacities.

Subsystems for 12 V supply in the vehicle and 24 V supply of body structures

In many cases it is necessary to make available a 24 V power supply for body-structure manufacturers in vehicles with 12 V electrical systems. This can be solved by means of a 12 V/24 V DC/DC converter and if necessary an additional 24 V battery. A cheaper solution is to use two series-connected 12 V batteries, replace the 12 V alternator with a 24 V variant, and use a battery equalizer. The 12 V electrical system including starter is supplied via the 12 V mid-tap while the vehicle body is supplied via the 24 V tap.

Dual-battery electrical systems

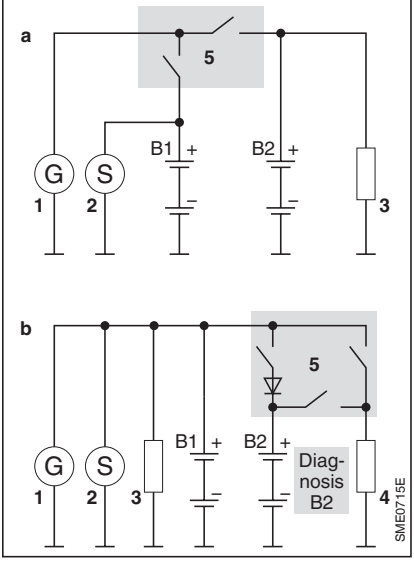
Separate supply for engine starting and recreational application

In long-distance vehicles in which drivers sleep overnight the batteries are cycled by frequent exhaustive discharging. This cycling causes starter batteries, which are optimized for high power output, to age quickly. In addition, to guarantee cold-starting performance, a charge reserve must always be maintained, meaning that only part of the full battery capacity can be used for the residential phase. A remedy is provided here by dual-battery electrical systems, in which a cold-start-optimized battery (e.g., a wet-cell battery) and a cycle-resistant battery for recreational application (e.g., a gel battery) are used. An energy-management function must control a switch arrangement in such a way that while driving both batteries are charged, when the engine is switched off the consumers are supplied only from the cycle-proof battery, and during an engine-start request the cold-start-optimized battery supplies the starter (Figure 2a).

Redundant supply of safety-relevant consumers

In many vehicles a redundant power supply is required for safety-relevant consumers, e.g., for electric steering. A simple way of achieving this is to install a second, for

Figure 2: Dual-battery electrical systems
a) Separation of batteries for recreational application and starting,
b) Redundant supply.
1 Alternator, 2 Starter, 3 Consumer,
4 Safety-relevant consumer,
5 Switch arrangement.
G Alternator, S Starter,
B1 Cold-start-optimized battery for engine starting,
B2 Cycle-resistant battery for recreational application (a) or redundant battery (b).



the most part much smaller battery with the same nominal voltage (12 V or 2×12 V in series), from which the safety-relevant component can be supplied over a limited period of time if the primary electrical system fails. A circuit in combination with a software function (Figure 2b) ensures that the redundant battery is charged in normal driving operation. The availability of the power reserve must be monitored by means of diagnostic functions and, if the primary electrical system fails, the supply to the safety-relevant component must be switched over by the redundant battery.

Truck-specific implementation of electrical systems

Ground return in truck electrical systems

The return of the negative cable (ground cable) cannot be effected via a consistently conductive chassis in trucks on account of the frame structure with mounted cab. Instead, ground return is effected via a combination of conductive vehicle components such as engine block, clutch or transmission housing, and by a separate ground cable.

Some vehicle manufacturers use the vehicle frame as a ground conductor while other manufacturers lay separate ground cables along the frame. The decision as to the optimal solution depends among other things on the vehicle body and the vehicle manufacture. It is in part technically complex to create a long-lasting low-resistance connection between painted frame parts. Another benefit of using separate ground cables is the monitoring of mass current via current distribution. The connection of conductive vehicle components and ground cables is established by welded-on or pressed-in stud bolts to which ground cables are screwed with cable lugs.

Ground distribution is effected in a star arrangement and the respective ground subsegments are brought together at ground distributors. Such ground distributors can be for example plates with pressed-in stud bolts secure to the engine block or clutch housing.

Connection of body structures and trailers

The power supply to the trailer is provided by the trailer socket, which – protected by one or more fuses – is connected to one of the vehicle power distributors. The connections for vehicle bodies are likewise tapped at vehicle power distributors.

If only small currents are needed for body structures, these can also be supplied via a power sub-distributor or a body-structure-manufacturer interface ECU.

Body structures are typically connected not by the vehicle manufacturer but after the basic vehicle has been completed by a body-structure manufacturer. To ensure the integration of the electrical supply of the body structure in the basic electrical system, it is necessary to take into account typical supply scenarios in the design of the basic electrical system. In addition, the interfaces and specifications to be observed in body-structure-manufacturer guidelines must be described in detail.

Batteries for truck electrical systems

Essentially, the 24 V and 12 V batteries in trucks are subject to similar requirements to those imposed on passenger-car batteries. Truck-specific requirements are derived from the cycle resistance and – as a result of the installation location on the minimally damped frame – from the vibration resistance. At present only lead-acid batteries are used for the 24 V or 12 V electrical system in trucks. Due to the high cost pressure, wet-cell batteries are still frequently used as starter batteries in trucks, but newer technologies such as AGM batteries or gel batteries are also used as optional extras. Typical battery capacities for 24 V electrical systems are 100 to 140 Ah for light-duty trucks and 140 to 225 Ah for heavy-duty trucks.

Energy management for truck electrical systems

The energy management of truck electrical systems also essentially corresponds to that of passenger-car electrical systems. Because in the case of trucks the focus is firmly on vehicle availability and total costs, the design objective of energy management is the optimum between fuel consumption, battery life, and vehi-

cle availability. Truck-specific adaptations of energy-management functions are typically made particularly with regard to load management (e.g., availability of an electric loading ramp, independent air conditioner), intelligent alternator activation (large variety of vehicle weights and thus the distribution of pulling and overrun phases, variety of usage profiles, highly reliable guarantee of battery life and cold-starting performance) or the actuation of isolating switches.

Overview of consumers

As is already known from passenger-car electrical systems, electrical consumers are essentially distinguished on the basis of their switch-on duration (continuous, long-term, and short-term loads). The switch-on durations are very much dependent on for example the time of year and the vehicle's operating mode. In the case of long-distance commercial vehicles there is the additional mode of sleeping in the vehicle overnight (recreational phase). Table 1 summarizes the average electrical-system load in watts of a typical long-distance truck in the various operating modes.

The design-relevant load cases are especially important in the dimensioning of the electrical system. These load cases differ, depending on the design criterion (e.g., charge balance, voltage stability).

In addition to the charge balance, it is also necessary to safeguard the voltage stability and voltage quality in the vehicle electrical system. The voltage limits of the consumers and the different load cases are available in various standards (e.g., LV124 [1], VDA320 [2], SAE J1455 [3], ISO 16750-2 [4]) and differ, depending on the manufacturer, marker or application.

Table 1: Average electrical-system load of a typical long-distance truck

Average electrical-system load [W]	Summer	Winter (0 to 15 min)	Winter (>15 min)	Recreational (>8 h)
Drivetrain including exhaust-gas treatment	171	386	218	
Chassis and driver assistance	53	149	53	
Exterior lights	207	207	207	
HMI (Human Machine Interface)	39	39	39	
Trailer	63	63	63	
Heating and air conditioner	40	1036	82	
Infrastructure Comfort&convenience and miscellaneous	107	107	107	288
Sum	680	1987	769	288

Electrical-system concepts for highly and fully automated driving functions

New power-supply requirements

The introduction of automated driving functions and driverless commercial vehicles is accompanied by an increase in the requirements imposed on the vehicle electrical system. At present there are still no certification regulations or binding standards; the state of the art continues to develop rapidly. The power-supply requirements are not yet fixed.

One source of requirements is SAE J3016 [5]. This defines and classifies automation systems for on-road motor vehicles in different levels (SAE levels 1 to 5). In this connection the standard distinguishes on the one hand between driver-assistance functions (SAE level 1 and level 2) in which the driver monitors the driving task and must be available as a fallback level. These driving functions support but do not replace the driver. Therefore there are no higher power-supply requirements here.

The situation is different for higher automation levels (SAE levels 3 to 5). In the case of these automation levels, after a fault in the vehicle – also including faults in the power supply – the automated driving function must be maintained in a fallback mode at least for a limited period of time until either the driver takes over the driving task or the vehicle is automatically placed in a safe state. Derived from this is the requirement that a fault in the power supply must not lead to the total failure of the sensors, control and actuators of the automated driving function.

Another source of requirements relating to the vehicle are the ECE regulations for the type approval of motor vehicles. Even if no concrete provisions concerning the power supply of highly and fully automated motor vehicles are defined in the 2020 version, requirements concerning the supply of power brakes can however be carried over to automated driving functions. The background is the intention of ECE R13 [6] that a single fault must not lead to an unbraked vehicle. If the braking system is designed in such a way that the driver cannot achieve a minimum deceleration of the vehicle with their physical strength, ECE R13 calls for separate monitoring devices and energy accumulators to provide the auxiliary energy needed for deceleration (for example with compressed-air brakes).

This requirement can be similarly carried over to the case of automated driving functions: If no driver is available or the driver does not have the task of being available as an immediate fallback level, two separate energy accumulators and two separate monitoring devices should be available in order also to bring a highly or fully automated vehicle safely to a stop after a fault in the power supply.

Voltage and energy requirements

The required amount of energy in the event of a fault strongly depends on which components in the fallback mode must be supplied over which period of time to guarantee the attainment of a safe state. This is dependent in particular on the configuration of the automated driving function and its area of application (Operational Design Domain, ODD). Thus the amounts of energy to be reserved are much smaller when the vehicle is to be brought to an immediate stop in the event of a fault than when much longer and more complex driving tasks are to be planned and implemented. Especially with regard to the requirement for a “mission completed” it can be much more advantageous also to provide the energy source by way of two channels.

Power supply for safety-relevant consumers

Various safety measures are possible in order to avoid a complete power-supply failure and to guarantee a "secure" power supply to achieve a safe state [7].

Fault avoidance

To avoid failures in the vehicle electrical system, it is necessary to implement for example development processes in accordance with the necessary safety rating (e.g., ASIL level as per ISO 26262 [8]).

Furthermore, it may be necessary to exchange components regularly and preventively, particularly if their typical service life is clearly below the specified vehicle service life, as is often the case with batteries.

Implementation of fault tolerance

Another possible way of maintaining the power supply for safety-relevant components is fault tolerance. The aim is, despite the presence of a fault, to maintain a (possibly degraded) automated driving function.

On the component level this means that for example internal faults such as short circuits must be internally encapsulated before they can significantly influence the voltage in the network. Similarly, energy sources must shut down as soon as they supply under uncontrolled conditions more energy than can be drawn.

The power supply can furthermore be subdivided into subsystems to which functionally redundant components of the automated driving function are distributed. If faults are detected in a subsystem, this can be isolated from other subsystems and thereby prevent fault propagation and safeguard the supply to the remaining components.

In addition, it is important to report detected faults and to react to them through the energy-management system or the safety-relevant driving function.

Fault detection

In the case of redundant systems it is important that in particular non-operating subsystems be checked before the safety-relevant function is activated. It is therefore essential, for example when switching the vehicle on but also on a regular basis during operation, to check the function of the power-supply components and thereby discover latent faults. This applies for example to the above-mentioned electrical isolating switches, which should isolate spontaneously occurring faults from the rest of the vehicle electrical system.

Fault prediction

The prediction of faults can furthermore ensure that an automated driving function is terminated before the onset of an imminent failure. Fault prediction is effected for example through a process of comparing the aging behavior of a component with the previous load cycles or evaluating wear indicators. This is particularly relevant to mechanical or chemical components (alternators, batteries, wiring harness) which exhibit pronounced aging and/or wear behavior.

Solutions, alternatives, challenges

To fulfill the fault-tolerance requirements described above, separate energy accumulators and separate monitoring devices are required to supply safety-relevant consumers. Figure 3 shows a basic Topology. As shown here, the power is provided from the power source. This can be performed simply or redundantly. Depending on the drivetrain configuration, the power source can be, as shown in Figure 4, an alternator on the same voltage level as the consumers or a 48 V system or an HV (high-voltage) system comprising electrical machine, traction battery, and DC/DC converter.

To achieve fault encapsulation, the individual subsystems are connected by

couplers in Figure 3. These couplers can disconnect the subsystems in the event of a power-source overvoltage or a short circuit in base-load consumers. The safety-relevant sensors, actuators and vehicle computers which are required for automated driving or to achieve a safe state can therefore be supplied from energy accumulators B1 and B2.

Couplers too come in different designs, as shown by way of example in Figure 4. These can be designed for example in the form of a DC/DC converter, an electronic semiconductor switch or an electronic power distributor like “Powernet Guardian”. It is also possible to implement in the couplers diagnoses to monitor the wiring harness for supplying the safety-relevant consumers which monitor for example overrange line and connector resistances.

Both lithium-ion and lead batteries are conceivable even for the energy accumulators. Double-layer capacitors (DLCs) as storage devices on the low-voltage side are a possibility only for a few applications with low power requirements on the fallback level.

It is fundamentally important for the power content and the performance of the batteries to be safely and reliably monitored by the electronic battery sensor (EBS) and/or the battery-management system (BMS). Here the voltage limits of the safety-relevant consumers are a significant criterion which must not be violated.

To ensure the absence of interaction between comfort and convenience consumers and safety-relevant components, the introduction also of electronic fuses in the main distribution or sub-distribution

Figure 3: Basic topology for automatic driving

1 Power source,
2 Base load, consumer group,
3, 4 Couplers,
5, 6 Group of safety-relevant consumers (sensors, vehicle computers, actuators).
B1, B2 Energy accumulators.

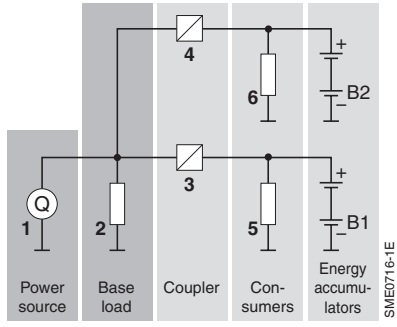
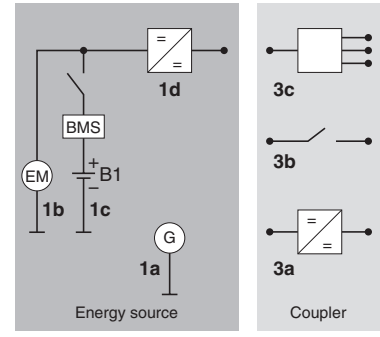


Figure 4: Different designs of the power source and couplers as a variation of the basic topology

1a Alternator,
1b Electrical machine,
1c Traction battery,
1d DC/DC converter,
3a DC/DC converter,
3b Electronic isolating switch,
3c Electronic power distributor.
B1 Energy accumulator,
BMS Battery-management system.



is in development. These can either be integrated in the established power distributors today or be part of the couplers or Powernet Guardian.

As mentioned above, there is the need to implement 12/24 V subsystems if complex passenger-car components like sensors, vehicle control computers or multimedia devices with 12 V voltage requirement must be integrated in the 24 V electrical system. If these fall into the group of safety-relevant consumers, this means that within the safety-relevant subsystems a voltage conversion must take place which further increases the complexity compared with passenger cars.

A second difference from a passenger car is that the supply to the truck trailer must also be taken into account. This contains if necessary safety-relevant consumer such as the ABS ECU, lighting or environment sensors, which depending on the automated driving function and its fallback level are required to attain a safe state. This also results, depending on the safety concept, in potentially the demand for redundancy of these functions.

This example shows that in the field of the power supply of automated driving trucks there are neither established requirements nor solutions and the state of the art will therefore only become established in the coming years.

References

[1] LV124: Electric and Electronic Components in Motor Vehicles Up To 3.5 t – General Requirements, Test Conditions and Tests.

[2] VDA320: Electric and Electronic Components in Motor Vehicles 48 V On-Board Power Supply.

[3] SAE J1455: Recommended Environmental Practices for Electronic Equipment – Design in Heavy-Duty Vehicle Applications

[4] ISO 16750-2: Road vehicles – Environmental conditions and testing for electrical and electronic equipment – Part 2: Electrical loads.

[5] SAE J3016: Taxonomy and Definitions for Terms Related to Driving Automation – Systems for On-Road Motor Vehicles.

[6] ECE R13: Regulation No. 13 of the Economic Commission for Europe of the United Nations (UN/ECE) – Uniform provisions concerning the approval of vehicles of categories M, N and O with regard to braking.

[7] Fail operational and ISO 26262, 2nd Edition; Carsten Gebauer, Robert Bosch GmbH, 11th international annual CTI Conference ISO 26262, September 2019.

[8] ISO 26262: Road vehicles – Functional safety.

12 V starter batteries

Requirements

The battery is used to start the internal-combustion engine – hence the designation starter battery – and to provide electrical energy for the 12 V vehicle electrical system if the alternator does not provide any energy or enough energy. Because the battery can be charged, it is in the proper sense a rechargeable battery.

The battery in the vehicle electrical system

The performance requirements for starter batteries in motor vehicles are very high. Diesel engines and large-volume gasoline engines have high cold-cranking power requirements with high starting currents, particularly at low temperatures.

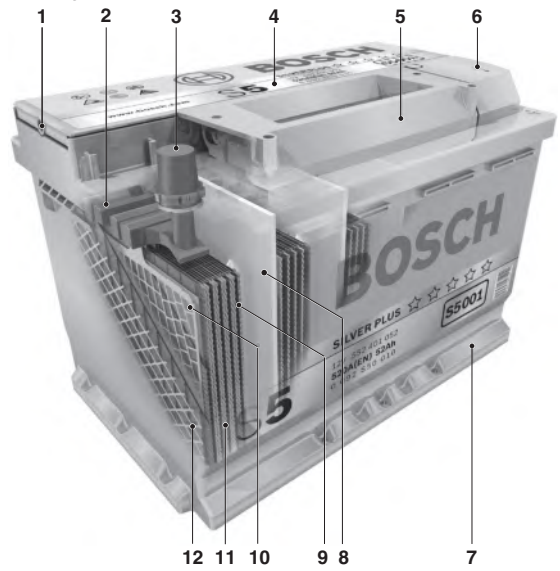
When the engine is running, the electric components are usually supplied directly from the alternator. When the engine speed is low, on the other hand, it is supplied partially from the battery.

The battery covers the temporary flow current consumption for ventilators, pumps and electronic components immediately after switching off the internal-combustion engine, as well as for comfort and convenience consumers operated while the vehicle is parked, such as entertainment and communication electronics and for the auxiliary heater, if present. If a vehicle is parked for days or weeks, a peak coil current of typically 3 to 30 milliamperes must be supplied. The battery must still be able to start the vehicle, even after it has been parked for extended periods. For seasonally operated vehicles the starter battery should be removed and connected to a trickle charger. This prevents the battery from suffering damage that can arise as a result of long inoperative periods in the exhaustively discharged state.

In addition to the uniform energy supply, the battery in the vehicle electrical system covers high dynamic current pulses which cannot be delivered by the alternator as quickly (for transient processes, such as for example switch-on processes in the electrical power steering). Furthermore,

Figure 1: Structure of a starter battery

1 Gas outlet opening,
2 Plate strap,
3 Battery terminal,
4 Block cover with labyrinth structure,
5 Carrying handle,
6 Terminal-post cover,
7 Bottom rail,
8 Cell partition,
9 Envelope separator,
10 Positive electrode, with stamped grid and positive active mass,
11 Electrode block (consisting of a set with positive electrodes with envelope separators and a set with negative electrodes),
12 Negative electrode, with expanded grid and negative active mass.



due to the very high inherent natural capacitance of the double-layer capacitor of a few farad, the battery smooths ripples in the automotive electrical system current, which helps minimize EMC problems (electromagnetic compatibility).

The nominal voltage of the vehicle electrical system is 14 V so that the 12 V battery can be charged with the alternator. With regard to the nominal voltage of the battery the vehicle electrical system is called the 12 V vehicle electrical system.

In commercial vehicles two structurally identical 12 V batteries are connected in series so that a higher starting power can be achieved for the starter with a 24 V vehicle electrical system. To avoid a nonuniform discharge of a battery, no additional 12 V consumers should be connected to a battery. This would lead to this battery "starving" because in the case of charging at constant voltage by the alternator the battery with the higher state of charge determines the charging current.

The battery in innovative vehicle electrical systems

Operating strategies of new vehicles use the battery in an increasingly active manner for reducing fuel consumption and emissions (see also 12 V vehicle electrical system). In vehicles with active alternator control the battery is operated in a partially charged state in order to be able to draw current from recuperation. The alternator output power is reduced in a targeted manner in operating phases (if necessary, all the load is taken off the alternator) in which the engine is working at poor efficiency or engine output is particularly necessary (e.g., during vehicle acceleration, "Passive Boost"). During these phases, the battery supplies the vehicle with electricity. In phases where the motor is highly efficient and when braking, the battery is then recharged in a targeted manner by raising the charge voltage (see Recuperation).

A start/stop system is used to switch off the engine if the vehicle is not moving (e.g., at a red light). The starter battery then covers the vehicle's entire power consumption of approximately 25 to 70 A. The engine is restarted automatically when the driver wants to pull away.

Depending on the operating conditions, using the battery's storage capacity in this way significantly reduces fuel consumption. The increased charging capacity represents an additional load on the battery, which is accommodated by special designs such as AGM and EFB (see Battery designs). Here, charging the battery well and sufficiently is always a requirement for a level charge balance in order to fulfill these tasks in the long term. A battery status detector is used to monitor the battery status. It generally consists of a sensor attached to the negative battery terminal (EBS, electronic battery sensor for measuring current, voltage and temperature) and software that uses the determined measurements to identify the status and capacity of the starter battery. Based on these results, the system can make the necessary intervention into the engine management system based on an operating strategy and, for example, decide whether the engine is to be shut off when the vehicle is stationary, whether it is to be restarted automatically when the vehicle is parked for long periods, or how the alternator voltage must be controlled (see 12 V vehicle electrical systems).

Numerous vehicles are also equipped with two batteries, for example, to guarantee startability even at a high standby current consumption or to prevent inconveniences (e.g., resetting the infotainment systems) due to the short-term voltage drop in vehicles with start-stop function when restarting automatically (see Two-battery vehicle electrical system).

Hybrid and electric vehicles also have a low-voltage vehicle electrical system of 14 V, on which the majority of the electric components work. This is supported by a 12 V battery, in many cases a high-performance starter battery. In some hybrid vehicles, cold cranking is carried out by a starter motor supplied from this battery.

The battery as component
The requirements mentioned define the electrical properties of the energy accumulator such as starting power, capacities and charge-current absorption. Furthermore, depending on operating conditions, thermal requirements (e.g., due to the installation point in the vehicle and the climate zone) and mechanical requirements (e.g., regarding attachments and vibration resistance) must be observed. The battery should also be maintenance-free, safe in use and environmentally friendly in production. The lead battery exhibits very good recyclability and achieves the highest recycling rates of all consumer durables at well over 95 %, where lead and plastic from old batteries are re-used for battery manufacturing.

The battery in operation
The design of the vehicle's components and the electrical operating concept, as well as the usage profile and driver affect the proper function of the battery and, therefore, the vehicle electrical system. Despite the excellent charge acceptance properties of modern starter batteries, a positive battery charge balance sometimes cannot be achieved during regular, short trips through the city in winter (involving high power consumption and low engine speed) and the state of charge decreases. At low states of charge, not only does the amount of energy still available decrease, but the capability of producing currents sufficient to start the engine decreases as well. Low battery charge levels over longer time periods worsen the cold-cranking characteristics and shorten the service life of the battery.

Battery design

Components
A 12 V starter battery has six electrode blocks which are installed in a polypropylene battery case subdivided by cell partitions (Figure 1, [1]). A cell consists of a stack of positive and negative electrodes arranged in alternation, each designed from a lead grid as a carrier and electrical arrester, filled with porous active mass (lead and lead dioxide, hence the designation lead battery). Microporous separators isolate the electrodes from each other, where they usually enclose the electrodes of one polarity in a pocket-size format. The separators are made from porous polyethylene filled with silica or from a glass-fiber fleece (AGM, Absorbent Glass Mat).

The electrolyte is in the form of diluted sulfuric acid which fills the pores in the electrodes and separators, and the voids in the cells. The electrodes of one polarity are connected in parallel by plate straps and, using cell connectors, are connected and sealed from openings in the partitions to the neighboring cells or the battery terminals. Battery terminals, cell and plate straps are made of lead alloys.

A lead battery develops, particularly during the charging process, a certain amount of oxygen at the positive electrode and hydrogen at the negative electrode. During this electrolysis process, water is consumed from the electrolytes. The charging gases must be released. The one-piece cover welded to the battery case provides the battery's upper seal and contains various ventilation openings depending on design. On conventional batteries, each cell has a plug, which is used for initially filling the battery with electrolyte, maintenance and the release of charging gases. Maintenance-free batteries usually no longer have a plug. Degassing occurs via an advanced labyrinth system, which prevents liquid from escaping, even if the battery is tilted. However, the ventilation openings, often arranged on the side of the cover, must not be sealed on both sides. Porous sintered bodies permeable to gas (called frits) are installed in front of the ventilation opening inside the cover. These sintered

bodies prevent potential flames or sparks outside from backfiring into the inside of the battery.

Conductive grid

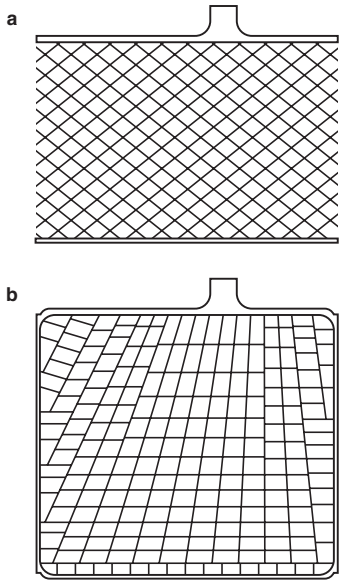
Active masses

Because electrochemical reactions in the battery take place very slowly, the electrodes are designed as grids to increase the size of the reaction surfaces. In the manufacturing process, the active masses are coated (pasted) into the conducted grid. The positive active mass contains after the initial charge (forming) porous lead dioxide (PbO_2, dark brown in color). The negative active mass contains porous metallic lead sponge (Pb, metallic gray-green in color).

Manufacturing and geometry

The lead grid is used for the mechanical mount and electric contact of the active masses. Optimizing the grid structure in terms of electrical conductivity improves the utilization of the active masses. The standard grid-manufacturing methods include casting liquid lead into a mold or continuous processes like drawing out (Figure 2a) and punching from a strip (Figure 2b).

Figure 2: Conductive grid
a) Expanded metal grid,
b) Stamped grid.



Alloy materials

Alloying elements are added to the lead used for the grid to optimize the production process and increase the mechanical strength and corrosion resistance. The strong oxidizing electrochemical potential exposes the positive grid to a continuous corrosion attack. As a result, the cross-sectional area of the webs detaches throughout the duration of use, the electrical resistance is increased and alloying elements are in released the electrolytes. Negative grids, on the other hand, do not have corrosive potential and, therefore, are not affected by corrosion (see Corrosion).

Antimony

If, for example, antimony assumes the function of the hardener for the positive grid, it is gradually released by corrosion and travels diagonally through the electrolyte and the separator to the negative electrode. It “poisons” the negative active mass, which greatly increases the spontaneous development of hydrogen there. This increases the self-discharge of the negative electrodes and water consumption if overcharged. Overall, this causes the capacity to be constantly reduced throughout the duration of use. The battery does not achieve the required state of charge, and the electrolyte has to be checked at frequent intervals. This is why, for grids, lead-antimony alloys (PbSb) with a low percentage of antimony by mass are still used only for battery types with plugs for which regular maintenance (replenishment of water) is accepted.

Calcium, tin and silver

Today, calcium, tin and silver are usually standard alloying elements for maintenance-free batteries, because they do not noticeably affect hydrogen development, even when released due to corrosion. To ensure the required strength, the terminals are made from lead alloys with up to around 10 % antimony by mass.

Lead-calcium alloys (PbCa) with around 0.1 % by mass are usually used for negative grids. To ensure the required strength, terminal posts are made from lead alloys with up to around 10 % antimony by mass.

Lead-calcium alloys (PbCa) with around 0.1 % calcium by mass are used for positive grids, which are manufactured in a continuous expanding or rolling and stamping process. Due to the increased percentage of tin (0.5 to 2 %), these alloys feature very high corrosion resistance, which allow lower grid weights.

Casted positive grids are often made from a lead-calcium-silver alloy (PbCaAg). In addition to 0.06 % calcium and tin by mass, this alloy also contains a proportion of silver (Ag). It has a finer grid structure and has proved to be extremely durable even at high temperatures – which have the effect of accelerating corrosion.

Table 1: Electrolyte values of the dilute sulfuric acid in a typical auto starter battery at 20 °C

State of charge	Electrolyte density in kg/*l*	Freezing threshold in °C
Charged	1.28	–68
Semi-charged	1.16…1.20	–17…–27
Discharged	1.04…1.12	–13…–11

Charging and discharging

Chemical reactions

The active masses of the charged lead battery are the lead dioxide (PbO_2) of the positive electrode, the sponge-like, highly porous lead (Pb) of the negative electrode, and the electrolyte (diluted sulfuric acid, H_2SO_4). The concentration in the charged battery is 37 %, producing an acid density of 1.28 kg/*l*. The electrolyte fills the pores of the active masses and the separator and is simultaneously an ionic conductor for charging and discharging. Compared with the electrolyte, PbO_2 and Pb assume characteristic voltages (individual potentials). Their difference is equal to the cell voltage of approx. 2 V at rest.

Specifically the processes shown in Figure 3 occur.

Discharged cell before charging

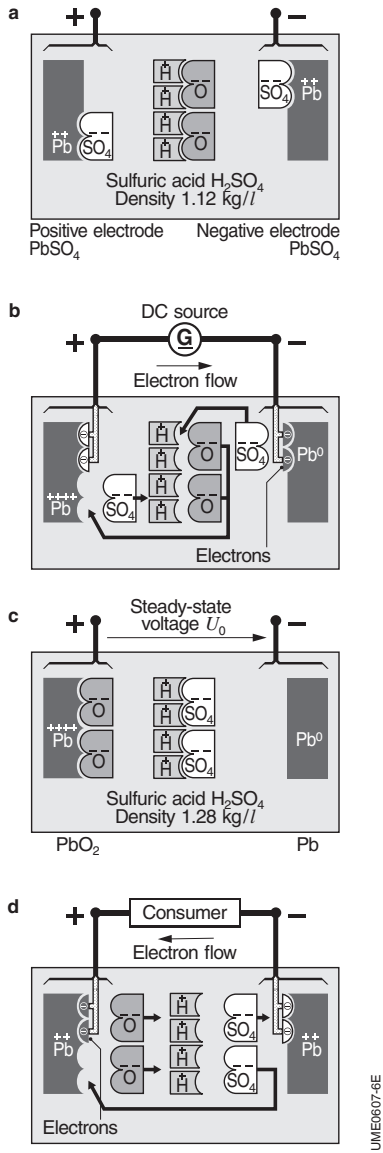
Located on both electrodes is $PbSO_4$, which is made up of the ions Pb^{2+} and SO_4^{2-} (Figure 3a). The battery acid has a lower density (approx. 1.12 kg/*l*, Table 1), as it is depleted of sulfate ions (SO_4^{2-}) by the previous current consumption.

Charging process

Pb^{2+} is converted at the positive electrode into Pb^{4+} due to electron donation (Figure 3b). This combines with oxygen arising from the splitting of H_2O to form PbO_2. On the other hand, elemental lead is formed at the negative electrode. Both these reactions involve the release of sulfate ions SO_4^{2-}, which with H^+ ions form sulfuric acid again and thereby increase the acid density.

The specific gravity of electrolyte can be used to indicate the state of charge of the battery (see Table 1). The accuracy of this relationship depends on battery design, electrolyte stratification, and battery wear with a certain degree of irreversible sulfating or a high degree of shedding of plate material. The electrolyte densities specified in Table 1 apply at a temperature of 20 °C; the electrolyte density drops by approximately 0.01 kg/*l* for every 14 K that the temperature rises and vice-versa when the temperature drops. The low value specified in Table 1 applies to high electrolyte utilization, the high value to low electrolyte utilization.

Figure 3: Charging and discharging processes in a lead battery
a) Discharged cell before charging,
b) Charging process,
c) Charged cell,
d) Discharging process.



Charged cell
$PbSO_4$ on the positive electrode is converted into PbO_2 and $PbSO_4$ on the negative electrode is converted into Pb (Figure 3c). There is no further increase in electrolyte density.

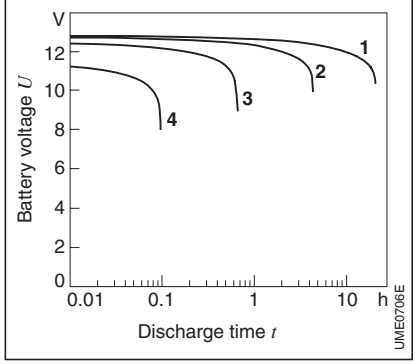
If the charge voltage continues to be applied after the cell has reached a state of full charge, only the electrolytic decomposition of water occurs. This produces oxyhydrogen gas (oxygen at the positive electrode, hydrogen at the negative electrode).

Discharging process
The direction of current flow and the electrochemical processes during discharging are reversed in relation to charging, which results in the Pb^{2+} and SO_4^{2-} ions on both electrodes being combined to form the discharge product $PbSO_4$ (Figure 3d).

When the cell discharges, PbO_2 and Pb combine with H_2SO_4 to form $PbSO_4$ (lead sulfate). This conversion causes the electrolyte to lose SO_4^{2-} ions (sulfate ions), and its electrolyte density decreases.

When a discharge current is applied to the battery, this reduces the battery voltage depending on the magnitude of current and the duration of discharge (Figure 4). Therefore, the charge quantity that

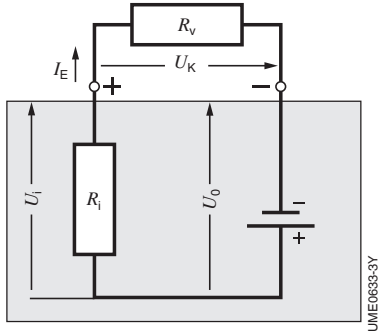
Figure 4: Battery-voltage curves as a function of the discharge duration for various discharge currents
1 I=0.05 A per Ah,
2 I=0.2 A per Ah,
3 I=1.0 A per Ah,
4 I=4.0 A per Ah.



can be drawn from the battery decreases when the current level increases.

Behavior at low temperatures
The chemical reactions in a battery always take place more slowly at lower temperatures and when internal resistance is increasing. The voltage at the specified discharge current and, therefore, the starting power of a battery even when fully charged decreases as the temperature drops and the internal resistance increases (Figure 5). The further the discharge progresses, the more the concentration of acid drops, up to the freezing point of the electrolyte. A discharged battery can only supply a low current which is not sufficient to start the vehicle.

Figure 5: Electrical parameters inside and on the battery
I_E Discharge current,
R_i Internal resistance,
R_V Load resistance,
U_0 Steady-state voltage,
U_K Terminal voltage,
U_i Voltage drop across internal resistance.



Battery characteristics

In addition to mechanical parameters, such as physical dimensions, attachment, and terminal design, characteristic electrical ratings are defined in testing standards (e.g., DIN EN 50342, [2], [3], [4]). General terminology on batteries is described in DIN 40729 [5].

For the cross-manufacturer marking of starter batteries, a nine-digit ETN (European Type Number) was described in the appendix of the revised EN 50342-1:2001 to compile the most important properties of a starter battery (including voltage, nominal capacity, cold cranking performance, terminal details, etc.). However, this representation cannot map expanded differences regarding construction characteristics, design and areas of application, which are described additionally in requirements outside of EN 50342-1. Therefore, starting with the version from 2006, the ETN nomenclature was withdrawn as a standard in coordination with battery manufacturers and users. Since then, the ETN only continues to be used occasionally.

Capacity
Capacitance is the charge quantity in ampere hours (A h) which can be drawn from the battery under specified conditions. It decreases as discharge current increases and temperature decreases.

Nominal capacity
EN 50342-1 defines the nominal capacity K_{20} as the charge the battery is able to deliver within 20 h up to a cutoff voltage of 10.5 V (1.75 V per cell) at a specified, constant discharge current $I_{20} = K_{20}/20$ at 25 °C. Nominal capacity is calculated from the quantities of active masses used (positive mass, negative mass, acid) and is not dependent on the number of electrodes.

Low-temperature test current
The low-temperature test current I_{CC} indicates the battery's current output capability at low temperatures. According to EN 50342-1, the battery terminal voltage with I_{CC} at –18 °C must be at least 7.5 V (1.25 V per cell) up to 10 seconds after discharging has started. Further details relating to the discharge period are specified in EN 50342. The short-term response of the battery when discharged at I_{CC} is largely determined by the number, geometric surface area of electrodes, the gap between electrodes, and the separator thickness and material.

The battery internal resistance R_i and other resistances in the starter circuit determine the cranking speed of the engine and also identify the starting response. For a charged battery (with 12 V), R_i at –18 °C is in the order of magnitude of $R_i \leq 4{,}000/I_{CC}$ (in mΩ), where I_{CC} must be applied in amps.

Water consumption
Lead batteries lose water from the electrolyte through electrolysis, particularly when charging. EN 50342-1 defines the test conditions for quantification. A limit value of water consumption of 1 g/A h, for example, means that, under test conditions, a battery with a 50 A h nominal capacity may lose up to 50 g of water. By comparison, the electrolyte of this kind of battery in new condition, depending on type, contains approximately 1.8 to 2.7 kg of water.

Cycle stability
Lead batteries undergo a certain amount of aging during repeated charging and discharging. Cycle stability means the capacity to be able to carry out a certain number of charging and discharging cycles under test conditions until the criteria at the end of service life are reached. EN 50342-1 defines the test conditions for quantification.

Battery types

Low-maintenance and maintenance-free batteries according to EN
Whether batteries require maintenance by refilling with water depends primarily on the grid alloy used. Batteries with lead antimony alloy (conventional and low-maintenance batteries) are very durable in cyclical charging and discharging operations, but require frequent maintenance because of the high degree of water loss and, therefore, are only used in some commercial vehicles.

Longer maintenance intervals result from low-maintenance batteries with negative grids made from a lead-calcium alloy (PbCa) and positive grids from an antimony alloy (PbSb). However, due to the antimony content in the positive grid, these hybrid batteries also seldom meet the high demand for very low water consumption (in accordance with EN 50342-1, less than 1 g/A h). Batteries with a water consumption under 4 g/A h in accordance with EN 50342-1 are considered maintenance-free in accordance with the EN standard.

Completely maintenance-free battery
Both grids in completely maintenance-free batteries are made from the lead-calcium alloy. When the vehicle electrical system is operating normally, this water decomposition is reduced to such an extent (in accordance with EN 50342-1, less than 1 g/A h) that the electrolyte reserves will last for the entire service life of the battery. Therefore, completely maintenance-free batteries do not require electrolyte level monitoring and do not usually provide a facility for doing this. They have a safety labyrinth cover that prevents electrolyte from escaping even when the battery is at a strong slope and are tightly sealed with the exception of two ventilation openings.

Completely maintenance-free batteries are already filled with electrolyte in the manufacturer's plant and can be stored in a fully charged state for up to 18 months after delivery due to the very low self-discharge.

A "dry" storage without electrolyte occurs only for very few designs (particularly for motorcycles). The electrolyte is

first poured in from a delivered acid pack when commissioning in workshops or at dealers' locations.

AGM battery

AGM (Absorbent Glass Mat) batteries are often used for applications with high charging capacities. In these batteries, the electrolyte is bound in a microporous glass-fiber fleece, which, instead of the conventional separators, is located between the positive and negative electrodes (Figure 6).

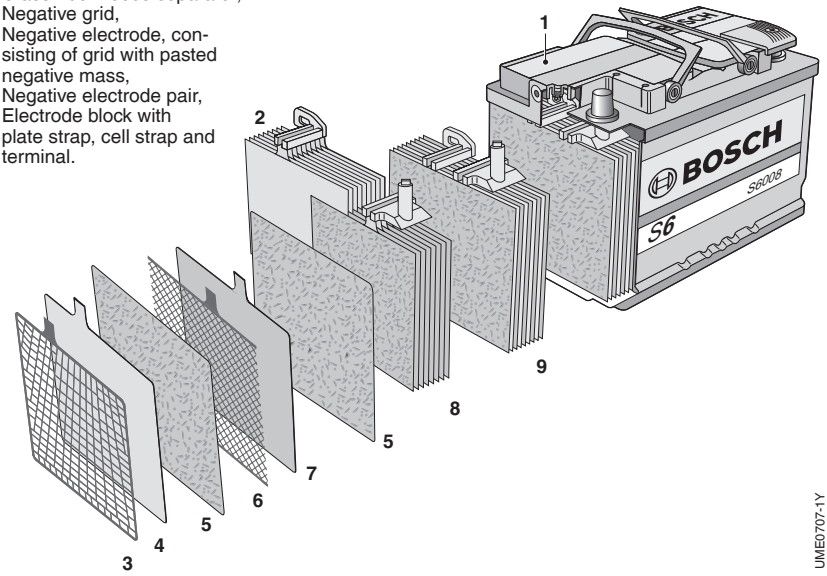
"Fixing" the electrolyte prevents, for example, the sulfuric acid of high concentration and density released during the charge from sinking and accumulating in the lower area of the cell. This "acid layer" with excess H_2SO_4 in the lower cell area and a shortage in the upper cell area occurs gradually during repetitive charging and discharging in conventional lead storage batteries with freely moving electrolyte. This impairs the charging behavior, accelerates the coarsening of the $PbSO_4$ discharge product's crystalline structure (sulfating), reduces the storage capacity, and accelerates the overall aging process of the battery.

In AGM batteries, these effects are reliably prevented. The elastic fleece also slightly pressurizes the electrode pair. This reduces the shedding and separation effect of the active mass considerably. Overall, this enables charging capacities more than three times higher than in comparable starter batteries with free electrolyte by charging and discharging throughout the entire service life. This makes it particularly well-suited for vehicles with a start/stop system that require the consumer to be supplied when the engine is stopped and then restarted reliably.

Other typical cases of application for AGM batteries include vehicles with many electrical consumers that are also used when the vehicle is stationary (e.g., auxiliary heater, consumer electronics), taxis, special vehicles and logistics vehicles.

Figure 6: AGM battery

1 One-piece cover with valves, central gas outlet openings and terminals,
2 Positive electrode pair,
3 Positive grid,
4 Positive electrode, consisting of grid with pasted positive mass,
5 Glass-fiber fleece separator,
6 Negative grid,
7 Negative electrode, consisting of grid with pasted negative mass,
8 Negative electrode pair,
9 Electrode block with plate strap, cell strap and terminal.



In the AGM battery, the oxygen that occurs at the positive electrode in an internal circuit is used up again, the amount of hydrogen that is produced there is suppressed and the amount of water loss is therefore minimized further. This circuit is enabled by small gas channels in the separator fleece, through which the oxygen is transported. The individual cells are separated from the environment by valves that open only if the operation causes the internal pressure to exceed a value of approximately 100 to 200 mbar. Under normal operating conditions, the valves are closed, which reduces water consumption even further. Lead batteries with sealed valves, such as the AGM battery, are called VRLA batteries (Valve-Regulated Lead-Acid Battery) and are completely maintenance-free.

The inner resistance is particularly low and high cold start currents are achieved because of the high porosity of the glass-fiber fleece. Therefore, AGM batteries are also often used in diesel vehicles.

Even if the battery housing is destroyed (e.g., in case of an accident), the electrolyte bound in the glass-fiber fleece (diluted sulfuric acid) does not generally leak. This and the improved gassing characteristics make the AGM battery particularly well-suited for installation points in the passenger cell.

Gel battery

In an AGM battery the electrolyte is fixed by an absorbent glass mat. Another method for fixing the electrolyte is to add silicic acid to the liquid. The gel battery therefore has in principle the same design as a lead battery with liquid electrolyte, which is immobilized in a gel by the addition of silicic acid. This delivers a leakproof, maintenance-free battery which is sealed to the outside by a valve (VRLA, Valve-Regulated Lead-Acid).

The benefit of this method lies in the simple design, which through support of the active masses results in increased cycle stability. The drawback is that ionic migration is hindered by the gel, which results in increased internal resistance. This type is therefore generally not used as a starter battery in passenger-car applications; it shows off its strengths when used as a general-purpose battery in two-battery vehicle electrical systems or in solar or camping applications.

It is important to ensure that an adapted charger that takes into account the slower power consumption is used.

EFB battery

The EFB battery (Enhanced Flooded Battery) is a battery with free electrolyte optimized for use in vehicles with a start/stop system. Although an acid layer is also formed when cyclic charging batteries without fixed electrolyte, such as the AGM battery. However, the positive plates coated with polyester scrim, which provide additional support of the active mass, and further design features produce a more robust design and a greater deep-cycle resistance in comparison with standard batteries.

EFB batteries are generally used in vehicles with a start/stop system and low consumer loads.

Commercial vehicle battery

Bosch also provides completely maintenance-free batteries for commercial vehicles. Specifically in commercial vehicle applications, these batteries offer a substantial cost advantage. These battery types have a special labyrinth cover, which guarantees leakproof performance and features central venting. Porous frits integrated into the outlet opening prevent flames or sparks outside from backfiring into the inside of the battery. There are special design variations for applications with extreme cycle or vibration stress.

The housing dimensions of commercial vehicle batteries are described in the standard EN 50342-4.

Deep-cycle resistant battery

Due to their design (relatively thin electrodes, lightweight separator material), starter batteries are less suitable for applications which involve repeated exhaustive discharging, as this results in heavy wear on the positive electrodes (mainly caused by a loosening and shedding of active mass). A deep-cycle resistant starter battery has separators with glass mats that support relatively thick plates containing positive mass and therefore prevent pre-

mature shedding. Service life measured in charge and discharge cycles is nearly twice as long as a standard battery.

Vibration-proof battery
In a vibration-proof battery, the electrode blocks are fixed to the battery case by means of cast resin or plastic to prevent any relative movement between the two components. In accordance with DIN EN 50342-1, this battery type must pass a 20-hour sinusoidal vibration test (at a frequency of 30 Hz) and be capable of withstanding a maximum acceleration of 6 g. Requirements are therefore about 10 times greater than for standard batteries.

Vibration-proof batteries are mainly used, for example, in commercial vehicles, construction machines or tractor vehicles.

Heavy-duty battery
A heavy-duty battery combines the attributes of the deep-cycle resistant and vibration-proof types. It is used in commercial vehicles which are subjected to extreme vibrations, and where cyclic discharge patterns are commonplace.

Battery for extended current output
This battery type shares the basic design of a deep-cycle resistant unit, but has thicker and fewer electrodes. Although no low-temperature test current is specified, the starting power lies well below (35 to 40 %) starter batteries of the same size. This battery type is used during extreme cyclic charges, sometimes for driving purposes as well.

Motorcycle battery
Batteries for use in two-wheelers are used in the same way as in passenger cars. A battery with a much lower capacity than its passenger-car counterpart is sufficient for motorcycle use. These batteries accordingly have smaller dimensions.

One advantage of gel and AGM batteries is that they can be used in any location on account of the fixed electrolyte.

These battery types are often delivered “dry and precharged”. Prior to use, these batteries must be filled with the acid provided. The manufacturer's instructions and precautionary measures for handling sulfuric acid must be observed!

Because motorcycles are often only operated in the summer months, it is recommended to connect the batteries to a trickle charger while they are out of service in order to avoid exhaustive discharge caused by standby-current consumers, which will shorten the battery service life.

There are in the meantime also motorcycle applications that use a lithium-ion-battery.

Lithium-ion starter battery
It stands to reason to make use of the features of lithium-ion technology such as energy and power density, service life and cycle stability for starter batteries as well. Such batteries already appear occasionally on the market. Above all the LiFePO4 system is used for safety reasons.

The battery system must contain the necessary control and monitoring electronics so that in can be used as a spare battery for conventional vehicle electrical systems. Details about the features and function of the electrochemical system are provided elsewhere (see Traction batteries).

Using the battery

Charging

In vehicle electric systems, the alternator regulator specifies the voltage. In terms of the battery, this corresponds to the IU charging method, where the battery charge current I is initially limited by the output of the alternator and then reduces automatically if the battery voltage reached the control value (Figure 7). Since the current consumption capacity of the battery decreases at low temperatures, the charge voltage in the vehicle is usually regulated depending on the temperature (Figure 8). At battery temperatures well under 0 °C, current is consumed only very slowly.

The IU charging method prevents damage due to overcharging, and makes sure that the battery has a long service life. Modern battery chargers operate according to similarly defined characteristic curves.

Older devices, on the other hand, still operate using constant current or a W-characteristic curve. In both cases, once the full state of charge has been reached, charging continues with only a slightly reduced current, or possibly a constant current. This leads to high water consumption and to subsequent corrosion of the positive grid. In particular, maintenance-free batteries can be damaged by this in the long term.

Discharging

Shortly after discharging at a continuous current begins, the voltage in the battery drops to a value which only changes relatively slowly if discharging continues. Only shortly before the end of the discharging process will the voltage collapse sharply due to exhaustion (d.h. complete electrochemical implementation) of one or more of the active components (positive or negative mass, electrolyte).

Self-discharge

Batteries discharge over a period of time if no electrical equipment is connected. Causes include the spontaneous development of oxygen at the positive electrode and hydrogen at the negative one, as well as the continuous slow corrosion of the positive grid. Batteries should therefore always be stored in well ventilated rooms and sparking and naked flames must be avoided in the proximity of batteries.

Completely maintenance-free batteries with lead calcium grids have a self-discharge of approximately 3 % per month at 25 °C, which remains roughly constant over the entire service life and nearly doubles for each temperature increase of 10 degrees.

Batteries using the lead-antimony alloy lose approximately 4 to 8 % of their charge every month when new. As battery ages, this value can increase by up to 1 % or

Figure 7: Charging according to the IU characteristic
1 AGM battery,
2 Low-maintenance standard battery.

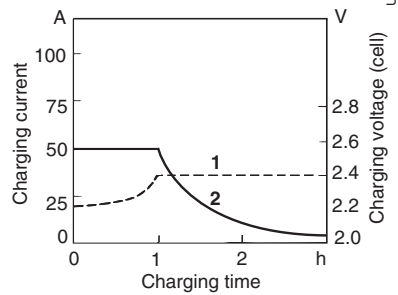
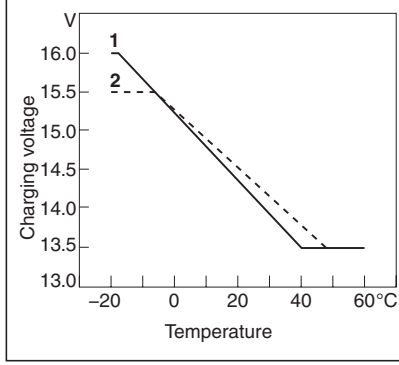


Figure 8: Temperature-dependent charging characteristic
1 Charge voltage,
2 Charging current.



more each day due to the migration of antimony to the negative electrode until a point is reached when the battery finally stops functioning.

Battery maintenance

On low-maintenance batteries, the electrolyte level should be inspected in accordance with the manufacturer's operating instructions; when required, it should be replenished to the MAX mark with distilled or demineralized water. This is generally not required for maintenance-free batteries in accordance with EN standard. This step is omitted entirely for completely maintenance-free batteries.

All batteries should be kept clean and dry to minimize self-discharge. Terminal posts, terminal clamps for connection with the vehicle electrical system, and fixing parts should be coated with acid-protection grease.

Before the onset of winter, it is recommend to check the battery status with a modern battery tester. If this tester recommends recharging, the battery should be recharged at a maximum voltage of approximately 14.4 to 14.8 V with a suitable charger with a regulated charging characteristic curve (IU charging characteristic curve or similar to prevent overcharging). Here, note the information in the user manual for the vehicle, battery and charging device. The charging area should be well ventilated (risk of explosion due to oxyhydrogen gas, risk of explosion, no naked flames or sparks). The electrolyte is corrosive. Therefore, gloves and protective eyewear must be worn when handling.

If no battery tester is available, a measurement of the electrolyte density or – if not possible – the steady-state voltage as a replacement. The battery should be recharged as described above if the electrolyte density is under approximately 1.24 g/m*l* or the steady-state voltage is under approximately 12.5 V.

Batteries temporarily removed from service should be stored in a cool, dry place. If the above-mentioned criteria exist, recharging is also required. Long inoperative periods at a low state of charge increase grid corrosion and reinforce sulfating. During this, the finely crystalline lead sulphate transforms into roughly crystalline, which makes the battery more difficult to recharge. These effects gradually damage the battery and can lead to their failure in the long term.

Battery malfunctions

Functional disturbances due to damage inside the battery cannot be remedied through repair. The battery has to be replaced.

A battery failure usually does not happen abruptly; instead, steadily decreasing starting power can provide an indication of the problem. Causes usually include a combination of consumed or sulfated active masses, grid corrosion and short-circuits due to separator wear. Modern battery testers can be used to evaluate whether a weak battery needs to be replaced or can simply be recharged.

If no battery defects are determined and the battery's charge is still always low, there may be a fault in the vehicle electrical system (defective alternator, electrical consumers remain switched on when the motor is not running, etc.). The battery can also be insufficiently charged as a result of using the vehicle primarily for short-distance driving. This can be balanced by occasionally driving longer distances or, alternatively, by occasionally recharging with an external charging device.

Battery testers

The starter battery is a wearing part. The service life is heavily dependent on the type and use of the vehicle and the climatic conditions. Aging becomes noticeable due to an increase in the internal resistance R_i (which impedes the start performance) and a reduction of the storage capacity by loosening of the active masses or sulfating. A low state of charge, on the other hand, is not a characteristic of aging, but a consequence of a usage with insufficient charge balance.

Often, a quick evaluation of the state of an older starter battery is desired, for example, to decide whether recharging is sufficient or a replacement is required. To do so, battery testing devices are available that can be connected to the powered-down battery and make a statement after a test duration of generally less than a minute.

Usually, these testing devices imprint a certain current charge profile onto the battery and deduce the battery state from the voltage response. To do so, they use partially simplified electrochemical impedance spectroscopy approaches (EIS, see electrochemical impedance spectroscopy). In general, the battery's correlation to the steady-state voltage is used to estimate the state of charge. Therefore, the battery must not be charged with other currents during the test.

Either a very rough evaluation (e.g., "ok", "charge", "replace", "test again") or a quantifying specification (e.g."state of charge 82 %", "startability 87 %") is given, depending on the design of the tester. Different manufacturers use a wide variety of methods.

Often, nominal values of the battery (capacities, cold cranking performance) must be specified and relative evaluations are made. The battery design (free electrolyte or AGM) and the battery temperature are also often requested or measured directly to also take into account this important influencing factor.

Battery testers cannot replace a complete test according to the applicable standards (e.g., EN 50342-1), but instead quickly give guideline for additional actions for batteries that have been in use for an extended amount of time. This test is not valid for new or like-new batteries because the structure of the active masses is often not yet balanced immediately after manufacturing.

Replacing the battery

Specifications in the vehicle's operating instructions must be taken into account when replacing batteries. These operating instructions often describe the specified or permissible sizes and design. The replacement battery should at least provide the specified capacity and cold cranking values to ensure functional and operational reliability. If a special battery design, such as AGM or EFB, is specified, such as in vehicles with a start/stop system, a replacement battery of the same design must be used.

Important: If the voltage in the vehicle electrical system drops to 0 V, consumers like the automotive sound system (car radio) may have to be restarted by entering a PIN. To avoid this situation, it is advisable when replacing the battery to connect a backup battery.

References

[1] A. Jossen, W. Weydanz: Moderne Akkumulatoren richtig einsetzen. 2nd Edition, Springer-Verlag, 2019.
[2] DIN EN 50342-1:2019-05: Lead-acid starter batteries – Part 1: General requirements and methods of test.
[3] DIN EN 50342-2:2015-10: Lead-acid starter batteries – Part 2: Dimensions of batteries and marking of terminals.
[4] DIN EN 50342-4: 2010-07: Lead-acid starter batteries – Part 4: Dimensions of batteries for commercial vehicles.
[5] DIN 40729:1985-05: Accumulators; Galvanic Secondary Cells; General Terms And Definitions.

Alternators

Electric power generation

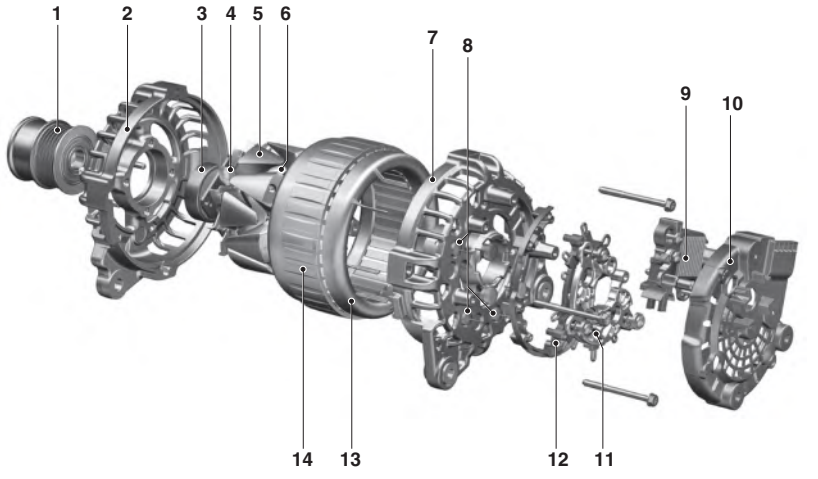
Motor vehicles need an alternator to charge the battery and to supply the electrical equipment such as ignition and fuel-injection systems, ECUs, lights, etc. with power. To charge the battery, it is necessary for the alternator to generate more current than is required by the electrical equipment when switched on. The alternator output, the battery capacity and the power demand of the electrical equipment must be matched to each other to ensure that under all operating conditions enough current is supplied to the vehicle electrical system and the battery is always sufficiently charged. This results in a consistent charge balance.
Because the battery and many items of electrical equipment have to be supplied with direct current, the alternating current generated by the alternator is rectified. The alternator is equipped with a voltage regulator so that it can supply the battery and the electrical equipment with a constant voltage.

Figure 1 shows an exploded view of an alternator. These alternators are designed for charge voltages of 14 V (in the 12 V vehicle electrical system) in passenger cars and of 28 V (in the 24 V vehicle electrical system) in many commercial vehicles.

Requirements

The essential alternator requirements are:
– maintenance of a direct-voltage supply to all electrical equipment in the system,
– charging reserves for (re)charging the battery, even at a constant load from electrical devices in continuous operation,
– maintenance of a constant alternator voltage throughout the entire ranges of engine speed and load conditions,

Figure 1: Exploded view of a (compact) alternator
1 Belt pulley with integrated freewheel or shock absorber,
2 A-side (drive-end) end shield, 3 A-side ball bearing, 4 A-side fan,
5 A-side claw pole, 6 B-side claw pole, 7 B-side end shield (collector-ring end shield),
8 Negative diodes (pressed in the B-side end shield),
9 Regulator assembly including brush holder and carbon brushes, 10 Protective cap,
11 Positive heat sink (with three holes for accommodating the positive diodes),
12 Rectifier wiring plate with connections for stator windings and diodes,
13 B-side winding overhang of stator, 14 Stator core.



– high level of efficiency,
– low operating noise,
– robust design capable of withstanding external stresses (such as vibration, high ambient temperatures, temperature cycles, dirt, moisture),
– long service life, comparable with that of the vehicle itself (for passenger cars)
– low weight,
– compact dimensions.

Principle of electromagnetic induction
To generate an alternating current a conductor loop is rotated between the north and south poles of a permanent magnet (Figure 2). If, as a result of the rotation of the conductor loop, the magnetic field in this conductor loop changes, a voltage is induced in it. According to Faraday's law of induction, the induced voltage U_{ind} increases as the speed v of the conductor-loop motion vertical to the magnetic-field lines increases and as the magnetic flux Φ passing through the conductor cross-section increases. The following applies:

$$U_{ind} \sim \frac{d\Phi}{dt} .$$

If the conductor loop consists not only of one turn but of n turns, this produces the n-fold induced voltage.

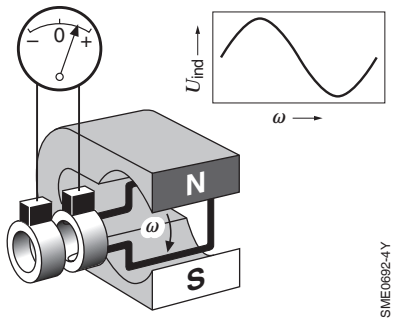
If the conductor loop rotates uniformly at the angular velocity ω the shape of the induced voltage is sinusoidal. The induced voltage can be tapped at the ends of the conductor loop via collector rings and carbon brushes. An alternating current flows when the circuit is completed.

The magnetic field for generating the induced voltage (excitation field) can be generated by permanent magnets. The advantage of these is that, thanks to the simplicity of their design, they do not have to be technically sophisticated. This solution is used in small alternators (e.g., bicycle dynamos). The disadvantage of excitation using permanent magnets is that it cannot be regulated.

A excitation field capable of being regulated can be generated with an electromagnet. An electromagnet consists of an iron core and a winding (excitation winding), through which an excitation current is passed. The number w of its windings determines together with the magnitude of the excitation current I_E the magnetomotive force θ ($\theta = w\, I_E$). The magnetomotive force of the excitation winding drives in the alternator's magnetic circuit – similarly to the voltage in an electric circuit – a magnetic flux Φ. The flux is directed in the closed magnetic circuit through highly magnetizable iron components and via air gaps as small as possible between the stator and the rotor. By altering the excitation current, it is possible to adjust the magnetic flux and with it the magnitude of the induced voltage.

When an external power source (e.g., a battery) supplies the excitation current, this is referred to as external excitation. When the excitation current is tapped directly from the generated alternator current inside the alternator, this is referred to as self-excitation.

Figure 2: Generation of an induced voltage in a coil permeated by a magnetic field
U_{ind} Induced voltage,
ω Angular velocity.



Alternator design

Claw-pole principle

The essential components of an alternator are the three- or multi-phase armature windings and the excitation system (Figures 1 and 3). Because the design of the armature-winding system is more complex than that of the excitation system and the currents generated in the armature winding are much greater than the excitation current, the armature windings are accommodated in the stationary stator – also called the frame (stator winding). The magnet poles with the excitation winding (rotor winding) are located on the rotating part, the rotor (inductor). The rotor's magnetic field is generated as soon as an excitation current passes through this winding. The alternator must feature a high number of pole pairs so that it can generate a high induced voltage already at low speeds. A high number of poles generates a high response per rotation and thus a high induced voltage. This is the prerequisite for a high alternator output.

By applying the claw-pole principle, the magnetic field of a single field coil can be split in such a way that the required 12 to 16 poles or 6 to 8 pole pairs are created (Figures 3 and 4). The number of poles than can be realized is limited. A low number of poles results in low machine utilization, while a high number of poles greatly increases the magnetic dispersion losses. For automotive alternators there is a good compromise from these two operative connections between 12 and 16 poles, depending on the output of the machines. Small alternators have smaller numbers of poles because the leakage fluxes increase with small pole pitches.

In the alternator, the armature contains three of more identical windings (phases) which are spatially offset in relation to one another (Figure 3). Because of the spatial offset of the windings, the sinusoidal alternating voltages generated in them are likewise out of phase with one another (temporally offset, Figure 5). The resulting alternating current is called three-phase alternating current or three-phase current.

Alternator circuits

In three-phase alternators, six current leads would be needed to transmit electrical energy in the case of non-connected windings. This number can be reduced to three by linking the three circuits. The circuits are linked in a star connection (Figure 6a) or in a delta connection (Figure 6b). The winding starts are usually designated u, v, w and the winding ends x, y, z.

In the case of a star connection, the ends of the three winding phases are interconnected at a single point, the star point. Without a neutral conductor the sum total of the three currents to the star point is zero at every moment.

Figure 3: Basic design of a claw-pole alternator with collector rings

1 Rotor,
2 Excitation winding (rotor winding),
3 Armature winding (stator winding),
4 Stator, 5 Collector rings, 6 Brushes,
7 Rectifier diodes.
B+ Battery positive terminal,
B– Battery negative terminal.

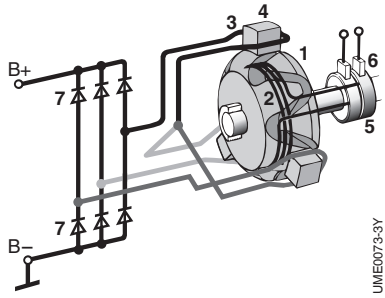
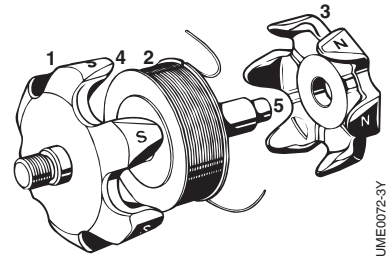


Figure 4: Components of a 12-pole claw-pole rotor

1 A-side (drive-end) claw pole,
2 Field coil,
3 B-side claw pole,
4 Chamfer on claw-pole-finger edge,
5 Alternator shaft.



Rectification of alternating voltage
The alternating voltage generated by the alternator must be rectified, since direct current is required to supply the battery and the electronics in the vehicle electrical system.

Two power diodes are connected to each phase – one diode on the positive side which is usually connected via the positive heat sink of the rectifier to the B+ stud and one diode on the negative side which is connected via the negative heat sink of the rectifier to the alternator casing (B– terminal) (Figure 3). The alternator casing is also electrically connected to vehicle ground via the alternator mounting points. The positive half-waves are conducted by the diodes on the positive side (B+ stud), the negative half-waves by the diodes on the negative side (vehicle ground). This principle is called full-wave rectification (Figure 5).

Full-wave rectification of the three phases with the so-called B6 bridge circuit brings about the addition of the positive and negative envelope curves of these half-waves to a rectified, slightly rippled alternator voltage.

The direct current which the alternator delivers under electrical load via terminals B+ and B– to the vehicle electrical system is not smooth, but slightly rippled. This ripple is further smoothed by the battery in parallel with the alternator and, if necessary, by capacitors in the vehicle electrical system.

In the case of alternators where the stator winding is star-connected (Figure 6a), two additional diodes (therefore also called "supplementary diodes") rectify the positive and negative half-wave of the star point. By rectifying the third harmonic in the phase voltages, these supplementary diodes can increase the alternator output current at 6000 rpm by up to 10 %. Alternator effi-

Figure 5: Three-phase alternating-current rectification
a) Three-phase alternating voltage,
b) Alternator voltage formed by the envelopes of the positive and negative half-waves,
c) Rectified alternator voltage.
U_P Phase voltage,
U_G Voltage at rectifier (negative not to ground),
U_{G-} Alternator direct voltage (negative to ground),
$U_{G,eff}$ Effective value of direct voltage.
U, V, W Strands.

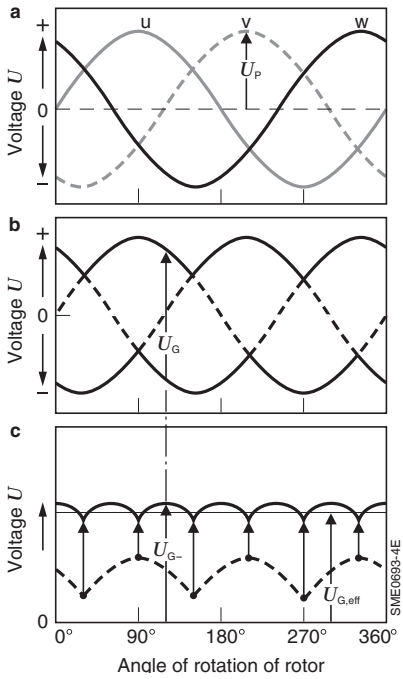
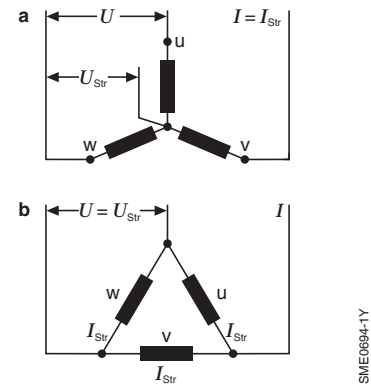


Figure 6: Connection types of three-phase stator windings
a) Star connection,
b) Delta connection.
U Alternator voltage, U_{Str} Phase voltage,
I Alternator current, I_{Str} Phase current.
U, V, W Strands.



ciency is thus significantly improved in the upper speed range. In the lower speed range, the amplitude of the third harmonic is below the vehicle system voltage such that the supplementary diodes cannot deliver any current. Supplementary diodes are rarely used in modern-day alternators.

Instead of high-blocking-capability power diodes, modern-day motor-vehicle alternators use Zener diodes to rectify the alternating voltage. Zener diodes limit voltage peaks that occur and thereby protect the alternator regulator and sensitive electrical-system consumers against overvoltages which can be caused for example by the deactivation of consumers (load-dump protection). When using a 14 V alternator, the response voltage of a rectifier fitted with Zener diodes ranges from 25 to 30 V.

Reverse-current block

The rectifier diodes in the alternator serve not only to rectify the alternator voltage, but also to prevent the battery from discharging via the stator windings. If the engine is stopped or if the engine revs are so low (e.g., starting speed) that the alternator is not yet excited, without diodes a battery current would flow through the stator winding. The diodes are polarized with regard to the battery voltage in the reverse direction so as to prevent the flow of any significant battery discharge current. Current can therefore only flow from the alternator to the battery.

Generally, alternators are not provided with polarity-reversal protection. Reversal of battery polarity (e.g., mixing up the battery poles when using an external battery to start the vehicle) can lead to destruction of the alternator diodes and poses a threat to the semiconductor components of other devices.

Alternator electric circuits

The standard version of an alternator has three electric circuits:

– pre-excitation circuit (external excitation by battery current),
– excitation circuit (self-excitation),
– main circuit.

Pre-excitation circuit

Before self-excitation of the alternator can start, a voltage must be induced in the stator which can drive the excitation current. The rotor does also have a low residual magnetism (remanence) in the unenergized state, but this is not sufficient for self-excited operation. The alternator must therefore be externally excited at the start of operation. This is done via the pre-excitation circuit, which is supplied by the battery.

Previously the pre-excitation current was directed via the charge-indicator lamp and if necessary to a resistor connected in parallel to the charge-indicator lamp. As soon as the voltage at the stator winding was high enough for the excitation diodes to have switched through, the excitation winding was supplied by the current generated in the alternator (self-excitation). Because the old regulators were switched through with the alternator stationary, the excitation circuit had to be disconnected from the main circuit so that the battery was not discharged by the excitation circuit with the vehicle at a standstill. Thanks to the "intelligent" multifunction regulators and interface regulators which have come to be used in the meantime, it has been possible to dispense with these excitation diodes because the new regulators deactivate the excitation current automatically with the vehicle at a standstill. When the multifunction regulator (see Voltage regulation) receives the information "ignition on" via the communication interface or connection L (Figure 7), it switches on the pre-excitation current with the permanently set on/off ratio (approximately 20 %, "controlled pre-excitation"). As soon as the rotor is turning, the regulator detects a voltage signal at its phase connection V, and from its frequency it can derive the alternator speed. When the switch-on speed set in the regulator is reached, the regulator switches through the output stage completely (on/off ratio 100 %) so that the alternator starts to supply current to the vehicle electrical system.

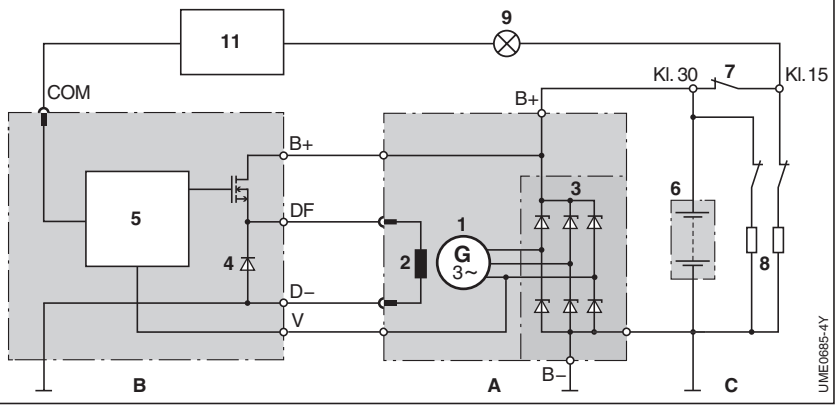
Excitation circuit
The purpose of the excitation current I_{err} is to generate a magnetic field in the rotor's excitation winding during the alternator's entire operating period and thereby induced in the stator windings the required alternator voltage. Since alternators are self-excited generators, the excitation current is branched from the stator winding. Alternators with multifunction regulators draw the excitation current directly from terminal B+ (Figure 7). The excitation current flows through the rectifier's power-plus diodes via the multifunction regulator, the carbon brushes, collector rings, and the rotor winding to ground (B–).

Main circuit
The alternating voltage induced in the alternator's phases must be rectified by the bridge circuit fitted with power diodes and directed to the battery and the electrical loads/consumers.

The alternator current I_G flows from the stator windings via the power diodes to the battery and to the consumers in the vehicle electrical system. The alternator current is divided into the battery charging current and into the consumer current.

Figure 7: Alternator
A Alternator, B Alternator regulator, C Vehicle electrical system.
1 Stator with armature windings,
2 Rotor with excitation winding (power supply via carbon-brush/collector-ring system),
3 Rectifier diodes, 4 Free-wheeling diode, 5 Regulator logic, 6 Battery,
7 Driving switch, 8 Consumers, 9 Alternator indicator lamp,
10 Relay (switches equipment which is to be switched on only when alternator is active),
11 ECU.
COM Communication interface (e.g., LIN bus),
DF Dynamo field, V Phase signal,
B+ Battery positive, B– Battery negative, D– Chassis ground.

Older alternator regulators without communication interfaces (now only customary in the commercial-vehicle sector) direct signals with the following designations to the plug connector:
W Digitized speed signal (alternator phase) for speed evaluation,
DFM (DF monitoring), S Sensing (battery-voltage supply lead),
L Lamp connection (activation of alternator indicator lamp).



Voltage regulation

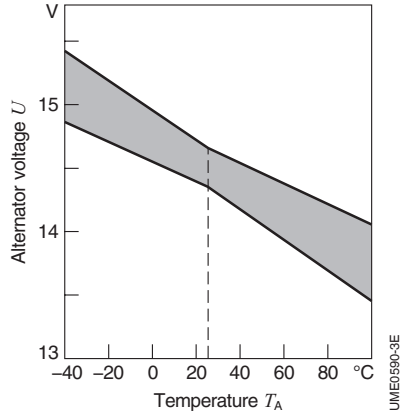
Function of voltage regulation

At constant excitation current, the alternator voltage is dependent on the alternator speed and load. The function of voltage regulation is to keep the alternator voltage – and thus also the vehicle system voltage – constant over the full speed range of the vehicle engine, irrespective of electrical load. To this end, the voltage regulator regulates the extent of the excitation current and consequently the magnitude of the magnetic field in the rotor depending on the voltage generated in the alternator. In this way, the regulator keeps the vehicle system voltage constant and prevents the battery from being overcharged or discharged during vehicle operation (see also Voltage regulation, 12 V vehicle electrical systems).

Vehicle electrical systems with 12 V battery voltage are regulated in the 14 V tolerance zone, those with 24 V battery voltage (commercial-vehicle sector) in the 28 V tolerance zone. As long as the voltage generated by the alternator is below the control voltage, the voltage regulator does not switch and the regulator output stage is switched on (on/off ratio 100 %).

The regulator characteristic curve (alternator voltage as a function of temperature) is adapted to the chemical properties of the battery. At low temperatures the alternator voltage is slightly higher to improve the battery charge in winter. The input voltages of the electronic devices are taken into consideration here. At higher temperatures the alternator voltage is lower so as to avoid battery overcharging and gassing in summer. An example of a characteristic curve is shown in Figure 8. The voltage level is 14.5 V with a gradient of –10 mV/K (temperature compensation).

Figure 8: Regulator characteristic curve
Permissible tolerance band of the alternator voltage (14 V) as a function of the alternator intake-air temperature T_A.



Principle of voltage regulation

If the voltage exceeds the upper setpoint value, the regulator switches off the output stage for activating the rotor winding. Driven by the inductance of the field coil, the excitation current initially flows on via the free-wheeling diode connected in parallel to the rotor winding; excitation becomes weaker and as a result the alternator voltage decreases. If now the alternator voltage drops below the lower setpoint value, the regulator switches on the excitation current again. Excitation increases and with it also the alternator voltage. If the voltage exceeds the upper setpoint value again, the regulation cycle begins afresh. Since the regulation cycles are in the range of milliseconds, the mean alternator-voltage value is well regulated to the specified characteristic curve (Figure 8).

The ratio of the respective on- and off-times is decisive for the magnitude of the mean excitation current. Because the performance of the alternator is lesser at low than at high rotational speeds, the excitation current must be adapted with changes in rotational speed. At low rotational speeds the on-time of the regulator output stage (power transistor) is relatively long and the off-time short. The excitation current is only interrupted by the regulator for a short time and consequently its mean value is high. Conversely, at high rotational speeds the on-time is short and the off-time long. A low excitation current flows.

If the excitation current were to be interrupted by the regulator output stage, a voltage peak would be created by the inductance of the excitation winding. A free-wheeling diode is connected in the regulator in parallel to the rotor winding to prevent the creation of such voltage peaks, which would destroy the regulator. It takes

over the excitation current at the point when the regulator output stage is switched off and ensures that the excitation current decays slowly until the regulator output stage is switched on again.

Regulators

Voltage regulators were formerly produced using discrete components. Today they incorporate hybrid or monolithic circuits (Figure 7). When monolithic technology is applied, the control and regulator IC, the power transistor, and the freewheeling diode are all accommodated on a single chip.

Multifunction regulators and interface regulators

Multifunction regulators and interface regulators perform special functions in addition to the task of voltage regulation described above. These regulators are today fitted in all new compact alternators. They draw the excitation current directly from B+ so that no excitation diodes are required.

The following additional functions can be integrated in these regulators:

– controlled pre-excitation,
– recognition of "alternator turning" through evaluation of the phase signal at connection V,
– emergency regulation in the event of interruption of the line at connection L or of the interface line (communication interface),
– overtemperature protection by limiting the excitation current when the limit temperature is exceeded,
– load-response functions.
– In the case of interface regulators, the control voltage can be adjusted via the communication interface.

For the "load-response function" (LR) a distinction is made between LR Driving and LR Start. LR Driving brings about a slow correction of the excitation current after an electrical-system load has been cut in so that the engine has sufficient time at idle to compensate the alternator's rising drive torque. LR Start means that the alternator is inactive for a specific period after starting so that the engine can settle to a stable idle.

The power loss of the alternator indicator lamp in the instrument panel is frequently too great and disruptive. It can be reduced for example by transferring to an LED display. Multifunction regulators permit the activation of both bulbs and LEDs as display elements.

Intelligent alternator regulation

Voltage regulators with digital interfaces (interface regulators) are the response to increased demands for greater mutual compatibility between engine-management and alternator-regulation systems. Various interfaces (e.g., bit-synchronous interface or LIN interface) are used here depending on the vehicle manufacturer. These interfaces are the means used to communicate information which in older regulators was output via individual connector pins.

The control voltage can be adjusted via the regulator interface. If a low control voltage is set, this results in low or no alternator power output because the voltage generated by the alternator is at the battery-voltage or lower. The power input and the torque input at the alternator shaft decrease to low levels. On the other hand, the alternator voltage can be stepped up via the interface so that the alternator charges the battery and supplies the electrical equipment in the vehicle electrical system.

Because the alternator converts mechanical energy into electrical energy, the power input at the alternator shaft increases as the electrical power output increases. The internal-combustion engine needs more fuel to cover the alternator's power requirement. However, if the alternator voltage is stepped up in overrun phases, the energy tapped by the alternator at the crankshaft does not cost any additional fuel, because overrun fuel cutoff stops the fuel supply. When the driver presses the accelerator pedal, the engine ECU steps down the alternator voltage again with the aid of the communication interface to the regulator in order to relieve the strain on the engine by the alternator torque. This process is also known as intelligent alternator regulation and recuperation.

Interface regulators allow a fine-tuning of load-response functions to engine operating state, optimization of torque patterns to reduce fuel consumption, and adjustment of charging voltage to improve the battery charge state.

Characteristic alternator values

Alternator performance

Automotive alternators function through the constant transmission ratio to the engine in a large rotational-speed range (1,600 to 20,000 rpm). Characteristic curves reflect the characteristic performance of an alternator. The alternator full-load characteristic refers here to a constant alternator voltage and to a defined ambient temperature.

Figure 9: Typical characteristic curve of an alternator

n_{BL} Idle-speed range (dependent on transmission ratio and idle speed of internal-combustion engine),
n_{max} Maximum speed,
n_0 Zero-ampere speed, n_N Rated speed,
I_L Current at engine idle, I_{max} Maximum current,
I_N Rated current.

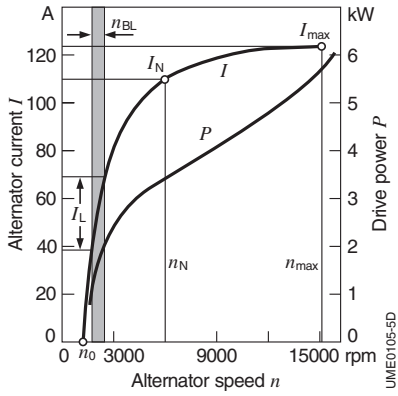
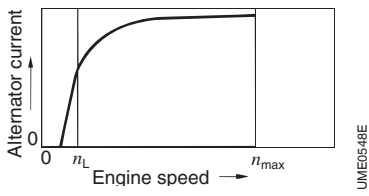


Figure 10: Characteristic curve of maximum alternator current at constant voltage

n_L Idle speed,
n_{max} Maximum speed.



Current characteristic

Defined rotational-speed points of the alternator are considered to describe the current characteristic (Figure 9).

Zero-ampere speed

The alternator can only deliver current to the vehicle electrical system when the induced voltage U_{ind} in the stator windings is greater than the sum total of the two diode forward voltages and the vehicle system voltage applied between B+ and B– (from approx. 1,000 rpm). The induced voltage increases with the frequency of the change in flux, i.e., with the speed of the rotor:

$$U_{ind} \sim \frac{d\Phi}{dt}.$$

There is a speed range between zero and the "zero-ampere speed", in which the alternator does not output any current (Figures 9 and 10) because the induced voltage in this range is too low.

Rotational speed and current at engine idle

The alternator speed at engine idle depends on the alternator's transmission ratio in relation to the engine. For compact-diode alternators this is typically in the range of n_L = 1,500 rpm, for compact alternators according to the usually higher transmission ratio typically in the range of n_L = 1,800 rpm.

The current generated by the alternator at engine idle speed (idle current) must at least be enough to supply the permanently switched-on loads/consumers.

Rated speed and rated current

The full-load current which the alternator outputs at the rated speed n_N = 6,000 rpm, at the alternator voltage U_G = 13.5 V, and at ambient temperature T_U = 23 °C is referred to as rated current.

Maximum speed and maximum current

The alternator current rises sharply under full-load conditions as the speed increases. The maximum current I_{max} is the highest current that can be generated by the alternator.

At higher speeds the voltage drop at the magnetizing reactance X_h (Figure 11) and the leakage reactance X_σ becomes so great that, in spite of an increasing

induced voltage, the current fails to rise significantly. The magnetizing reactance is obtained from that part of the coil inductance which generates the profitable induced voltage and whose flux takes the envisaged paths in the rotor and the stator. The leakage reactance forms that part of the flux which is not linked to the rotor – i.e., leakage fields which short, for example, in the end winding or directly via the groove.

The alternator's speed is first and foremost limited by its design (see Conditions of service).

Characteristic of drive power

The characteristic of drive power is used to design the drive belt. The characteristic shows by way of the engine speed range what power (maximum) is transmitted by the engine to the alternator. It is also possible to determine the alternator's efficiency from its drive power and power output. The characteristic of driver power exhibits a flat shape in the middle speed range and rises sharply at higher speeds.

Operating conditions

Rotational speed

The utilization of an alternator (generable energy per kilogram of mass) increases as the speed increases. It is therefore necessary to aim for as high a transmission ratio as possible between the crankshaft and the alternator. The alternator already outputs at least a third of its rated power when the engine is at idle speed. However, it is important to ensure that at maximum engine speed the maximum permissible alternator speed is not exceeded. Widening of the claw poles and the service life of the ball bearings used determine the maximum permissible speed of motor-vehicle alternators.

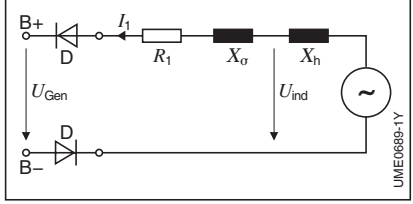
Typical values for the maximum speed are for compact alternators in the range of 18,000 to 22,000 rpm, for compact-diode alternators in the range of 15,000 to 18,000 rpm, and for alternators for commercial vehicles between 6,000 and 12,000 rpm. The transmission ratio for passenger cars is between 1:2.2 and 1:3, and for commercial vehicles up to 1:5.

Cooling

The losses arising during conversion from mechanical to electrical energy results in the alternator components being heated. The surrounding air in the engine compartment is used to cool the majority of the motor-vehicle alternators used today.

Figure 11: Simplified diagram of a phase with rectification

B+ Battery positive terminal,
B– Battery negative terminal,
D Rectifier diodes,
I_1 Alternator current,
R_1 Ohmic resistance,
X_σ Leakage reactance,
X_h Magnetizing reactance,
U_{ind} Induced voltage,
U_{Gen} Alternator voltage.



If the surrounding air is not sufficient to cool the alternator, fresh-air induction from cooler areas or liquid cooling are suitable measures for cooling the components adequately.

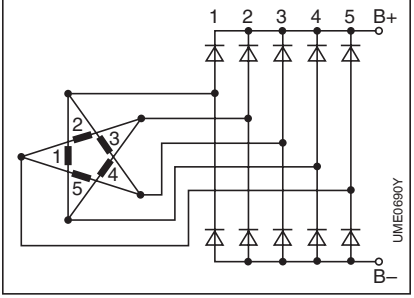
Engine vibrations
The alternator can, depending on the mounting conditions and the vibration characteristic of the engine, be exposed to vibration accelerations of 500 to 800 m/s^2. The mountings and the components of the alternator are thereby subjected to high forces. Critical natural frequencies in the alternator assembly must therefore be avoided at all costs.

Engine-compartment climate
The alternator is exposed to splash water, dirt, oil and fuel mist, and where necessary road salt (grit). To protect the alternator against corrosion, it is imperative to observe minimum distances between components with different electric potentials or to insulate the live components.

Acoustic behavior
The stringent demands placed on the noise emissions of modern-day vehicles and the superior smooth running of modern-day internal-combustion engines require quiet alternators. As well as chamfers on the claw-pole fingers which reduce the hard chopping of the magnetic flux at finger edges, there are other possibilities for reducing the noise magnetically generated by motor-vehicle alternators. The following may be used: firstly, five phases in a pentagram connection (Figure 12), and secondly, two three-phase systems which electrically offset by 30°.

Interference-suppression measures
The alternator and also other electrical loads/consumers of a motor vehicle can interfere with other electrical and electronic devices through their electromagnetic fields. Intensified interference suppression of alternators is then necessary if for example a radio installation, a car phone, or a car radio is operated in the immediate vicinity of or in the vehicle itself. Alternators are for this reason equipped with an suppression capacitor. On older compact-diode alternators a suppression capacitor can be subsequently mounted to the outside of the collector-ring end shield. On compact alternators it is already integrated in the rectifier or in the regulator housing.

Figure 12: Pentagram-connected five-phase stator winding with connected bridge rectifier as measure for improved acoustic behavior
B+ Battery positive terminal,
B– Battery negative terminal.



Efficiency

The ratio between the power which is supplied to the unit and the actual power output is known as the efficiency. Losses are an unavoidable byproduct of all processes in which mechanical or kinetic energy is converted into electrical energy. The losses in the claw-pole alternator are categorized as follows (Figure 13).

Losses in the claw-pole alternator

Copper losses in the stator and rotor windings

The ohmic losses in the stator windings and the rotor winding are called copper losses. They are proportional to the square of the current.

Iron losses in the laminated stator core

Iron losses result from hysteresis and eddy currents produced by alternating magnetic fields in the iron of the stator.

Eddy-current losses on the claw-pole surface

Eddy-current losses on the claw-pole surface are caused by fluctuations in flux brought about by the stator slotting.

Rectifier losses

Rectifier losses are caused by the voltage drop at the diodes. Standard diodes have a voltage drop of around 1 V. Rectifier losses can be reduced and thus efficiency improved by using semiconductors with a lower voltage drop, such as for example high-efficiency diodes with a voltage drop between 550 mV and 700 mV or active rectification diodes (ARDs), which with the aid of activated MOSFETs exhibit only a few millivolts of voltage drop.

Mechanical losses

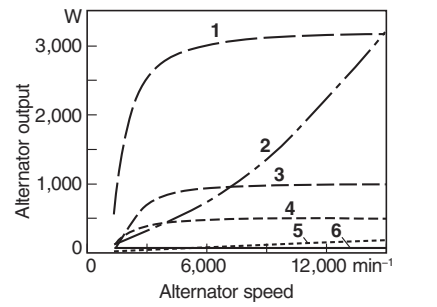
Mechanical losses include friction loss in the rolling bearings and at the sliding contacts, air resistance encountered by the rotor and, above all, fan resistance which rises dramatically as speed increases (aerodynamic losses).

Efficiency optimization

In regular automotive operation, the alternator operates in the part-load range. Efficiency at medium speeds is then 70 to 80 %. The use of a larger (and heavier) alternator allows it to operate in a more favorable part-load efficiency band for the same electrical load. The efficiency gains provided by the larger alternator more than compensate for losses in fuel economy associated with greater weight. However, the higher mass moment of inertia in the belt-drive system must be taken into consideration.

Figure 13: Loss distribution of a 220 A alternator

1 Alternator power output,
2 Iron losses,
3 Copper losses in the stator,
4 Rectifier losses,
5 Friction losses,
6 Copper losses in the rotor winding.



Alternator versions

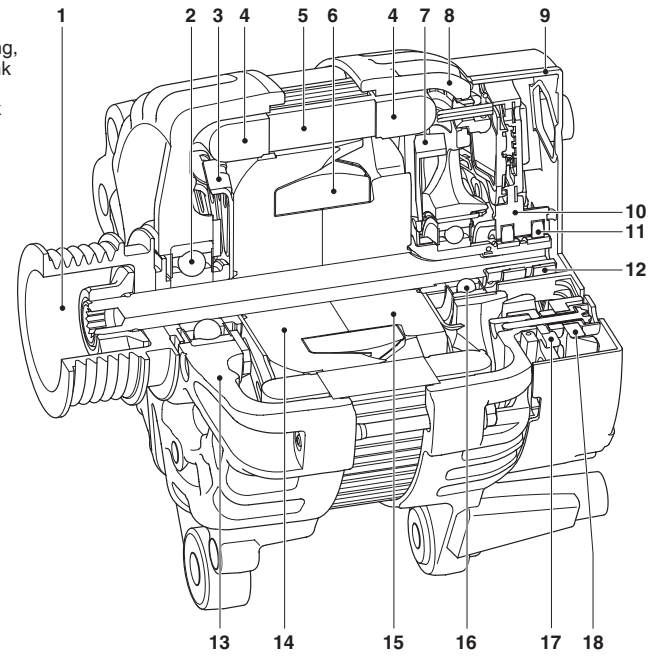
Claw-pole alternators have completely replaced the direct-current generators previously used as standard in motor vehicles. Based on equal outputs for both concepts, a claw-pole alternator weighs 50 % less, and is also less expensive to manufacture. Large-scale application (at the beginning of the 1960s) only became feasible with the availability of compact, powerful, inexpensive, and reliable silicone diodes.

Design criteria
Different basic designs have been developed for the various conditions of service and performance ranges of the respective vehicle types and their driving engines. The following criteria are crucial to the choice of alternators:

- vehicle type,
- operating conditions,
- rpm range of the respective internal-combustion engine,
- battery voltage of the vehicle electrical system,
- power demand of the loads/consumers,
- strain of the alternator by environmental influences (e.g., heat, dirt),
- installation conditions and dimensions.

Derived from these criteria is the required electrical dimensioning of the alternator, i.e.,
- alternator voltage (14 V or 28 V),
- maximum power output (voltage × current intensity),
- maximum current.

Figure 14: Compact alternator, EL Series (Efficiency Line)
1 Belt pulley, 2 A-side (drive-end) ball bearing, 3 A-side fan,
4 Stator winding overhangs, 5 Stator core, 6 Rotor winding (excitation winding),
7 B-side fan, 8 B-side end shield (collector-ring end shield),
9 Protective cap, 10 Brush holder,
11 Carbon brush, 12 Collector ring,
13 A-side end shield (drive end shield,
14 A-side claw pole,
15 B-side claw pole,
16 B-side ball bearing,
17 Negative heat sink of rectifier,
18 Positive heat sink of rectifier.



Design of an alternator for use in vehicles

The alternator casing consists of the drive end shield and the collector-ring end shield (Figure 14, see also Figure 1), between which the stator is mounted. The rotor shaft is mounted in the two bearing halves. The alternator stator consists of laminations provided with slots which are stacked and welded into a fixed laminated core. The slots contain the coils of the stator windings.

The rotor gives the claw-pole alternator its name. It consists of two opposingly polarized pole-wheel halves whose claw-like pole fingers mesh alternately as south and north poles (see Figures 3 and 4).

The two pole-wheel halves enclose the ring-coil-shaped excitation winding (see Figure 4). The excitation winding is supplied with current via the carbon-brush/collector-ring system (see Figure 3). Because the regulator adjusts the excitation current as a function of alternator voltage and set control voltage (nominal voltage), the brush holder with the carbon brushes is often integrated in the regulator housing. The regulator housing is mounted on the collector-ring end shield. Springs press the brushes against the collector rings so that, even under raw ambient conditions, they establish an engine a secure electrical contact with the collector rings on the rotor shaft.

The magnetic working flux passes through the pole core, the left pole half and its fingers, across the air gap to the fixed laminated stator core with the stator winding, and returns through the right pole-wheel half back to the pole core to complete the circuit.

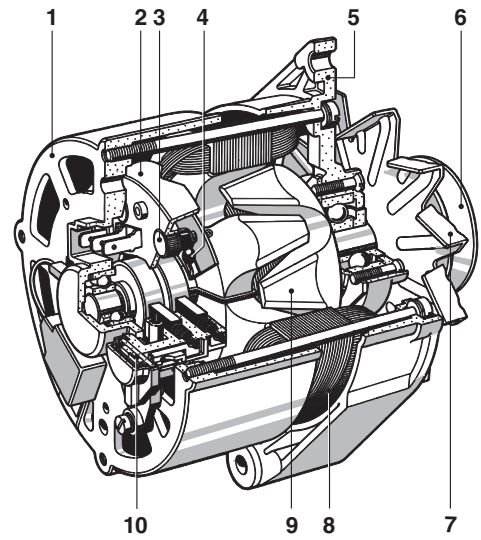
The electronic voltage regulator forms together with the brush holder a single unit if it is intended for mounting directly on the alternator. In some cases the electronic voltage regulator is mounted separately from the alternator in a protected place on the vehicle body or is integrated in a control unit (e.g., in the engine management). The external regulator can then be connected via electrical plug-in connections to the brush holder.

Compact alternator

The compact alternator is characterized by two internal fans which are mounted on the rotor at the drive and collector-ring ends. The fans draw in air in each case axially and discharge the heated air radially (double-flow ventilation). The two small

Figure 15: Compact-diode alternator

1 Collector-ring end shield,
2 Rectifier heat sink,
3 Power diode,
4 Excitation diode,
5 Drive end shield with mounting flanges,
6 Belt pulley,
7 External fan,
8 Stator,
9 Claw-pole rotor,
10 Transistor regulator.



fans produce substantially less aerodynamic noise than the large external fan of a compact-diode alternator. They are also suitable for higher rotational speeds (maximum speed: 18,000 to 22,000 rpm). These two properties provide for a higher transmission ratio between the crankshaft and the alternator so that compact alternators can output up to 25 % more power for the same engine speed and size.

The compact alternator is today used across the board in the passenger-car and commercial-vehicle sectors. The EL Series (Efficiency Line), the BL Series (Base Line) or the PL Series (Power Density Line) is used to suit the individual requirements. The service life of the alternator is a decisive factor in the commercial-vehicle sector. For this reason, larger ball bearings at the drive end and a brush/collector-ring system optimized for longer service life are used in the Classic Line (CL) and Heavy Duty (HD) Series and in the HDS (Heavy Duty Second Generation) Series.

The maximum permissible speed of commercial-vehicle alternators ranges between 6,000 rpm and 12,000 rpm.

Compact-diode alternator

The compact-diode claw-pole alternator (Figure 15) differs from the compact alternator first and foremost in the form of the large fan which is located between the belt pulley and the pot-shaped alternator casing. The external fan with a maximum speed of 12,000 to 18,000 rpm draws the cooling air axially through the casing (single-flow ventilation).

Because the rectifier, regulator, and brush/collector-ring system are arranged inside the end shield, the shaft inside the collector rings must be relatively thick to be able to transmit the belt-drive forces to the external ball bearing. The collector rings therefore have a large diameter, only permitting a limited brush service life.

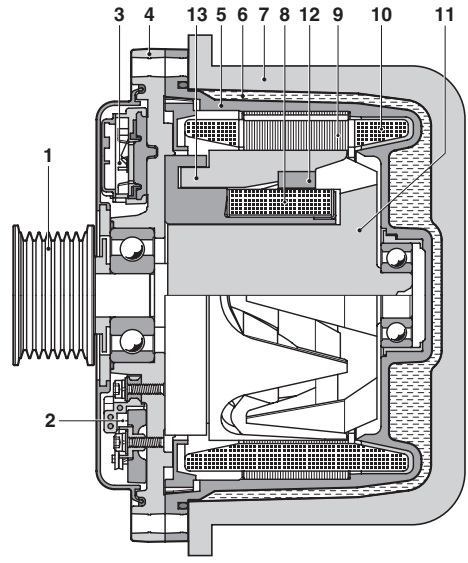
Compact-diode alternators can still be found in older vehicles.

Alternators with windingless rotors

The windingless-rotor alternator is a special design variant of the claw-pole alternator where only the claw poles rotate, while the rotor winding remains stationary. With this design, the regulator supplies directly to the rotor winding; sliding contacts are not required (Figure 16).

Figure 16: Liquid-cooled alternator with windingless rotor

1 Belt pulley,
2 Rectifier
3 Regulator,
4 Drive end shield,
5 Alternator casing,
6 Coolant jacket,
7 Jacket casing for engine mounting,
8 Stationary rotor winding,
9 Stator core,
10 Stator winding,
11 Windingless rotor,
12 Non-magnetic intermediate ring,
13 Conductive element.



The rotating parts consists merely of the rotor with pole wheel and conductive element. The conductive element consists as a rule of a magnetically highly permeable material (e.g., soft iron), which boosts the magnetic flux of a magnetic field acting on the conductive element or forwards it (e.g., to the stator core). Six pole fingers of the same polarity in each case form a pole-finger crown as north and south poles. Only one of the two pole-finger crowns is connected directly to the rotor shaft. A non-magnetic retaining ring, which is located below the meshing pole fingers, holds the two crowns together as claw-pole halves.

The magnetic flux passes from the pole core of the rotating rotor via the fixed internal pole to the conductive element, then across its pole finger to the fixed stator core. The magnetic circuit is completed in the rotor pole core via the opposingly polarized claw half. The magnetic flux must when compared with the collector-ring rotor overcome two additional air gaps between the rotating pole wheel and the fixed internal pole.

Thanks to the omission of the wear component collector-ring/carbon system, these alternators are particularly suitable for applications which require a long service life combined with heavy alternator load, e.g., for construction machinery or traction generators. The units weigh somewhat more than claw-pole alternators with collector rings due to the fact that additional iron and copper is required to conduct the magnetic flux through two additional air gaps.

Liquid-cooled alternator

The windingless rotor is also used in liquid-cooled alternators, among other applications. In the case of this alternator, engine coolant flows around the complete jacket and rear of the alternator casing. This significantly reduces noise when compared with air-cooled alternators in that the alternator is completely encapsulated and no fan produces aerodynamic noise. The electronic components are mounted on the drive end shield.

Current alternator series

Efficiency Line

Efficiency Line (EL) Series alternators are suitable for vehicle electrical systems with medium to high power demand. The Efficiency Line covers the rated-current range of 150 A through 250 A. As the name already intimates, Efficiency Line alternators convert energy with high efficiency. Five phases are used from a rated current of 180 A to improve acoustic behavior and reduce the voltage ripple. A winding with a high copper space factor (44 to 50 %) is pulled into the stator core.

Base Line

Base Line (BL) Series alternators were developed for vehicle electrical systems with low to medium power demand. Their rated current extends to 150 A. Owing to their compact design they can also be installed in the confined engine compartments of small vehicles. The stator core carries a three-phase pull-in winding with a copper space factor of between 40 % and 50 %.

Power Density Line

Power Density Line (PL) Series alternators were developed for vehicle electrical systems with medium to high power demand. The Power Density Line covers the rated-current range of 150 A through 250 A. The stator winding is inserted into a flat iron core. The flat core with winding is bent round and the ends are welded. The flat core delivers a copper space factor of over 60 %, which is crucial to the high power density of PL alternators.

All PL alternators are of five-phase design to keep noise emissions and voltage ripple as low as possible. The optimized cooling and high efficiency enable them to be operated at high ambient temperatures.

Starters for passenger cars and light commercial vehicles

Application

Overview
Internal-combustion engines require start assistance to start. They must be cranked by a starter at a minimum speed before they can supply sufficient power in sustained operation. A starter consists of the following assemblies (Figure 1):
– starter motor (electric motor, DC motor),
– reduction gear,
– engaging system with solenoid switch,
– overrunning clutch with starter pinion.

The starter torque is transmitted via the starter pinion and the engine-flywheel ring gear to the engine (see Figure 6). The starter pinion engages the ring gear during the starting operation. The ring gear with around 130 teeth is located on the engine flywheel or on the housing of the hydrodynamic torque converter. The starter pinion with around 13 teeth is situated a few millimeters in front of the ring gear in its rest position.

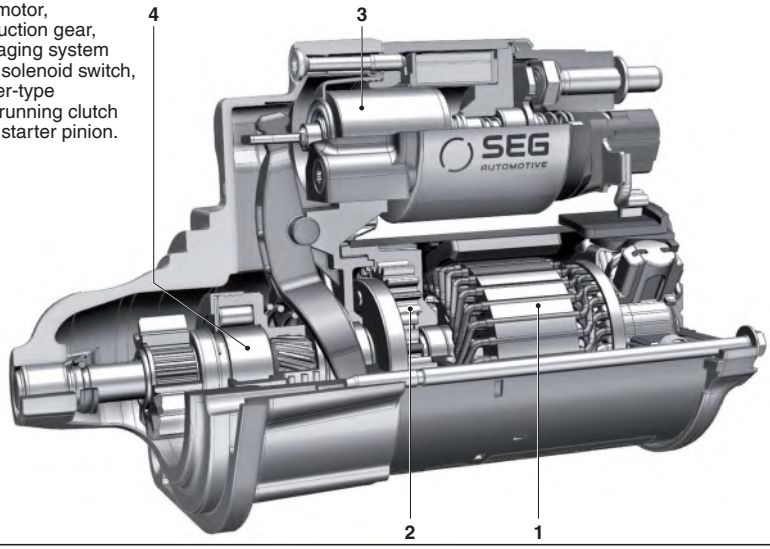
Requirements

Because of the high step-down ratio between the starter pinion and the ring gear, the starter may be designed for high speeds and low torque. This allows the dimensions and the weight of the starter to be kept small. The transmission ratio is in the range of around 1:10, enabling the speeds required for sustained internal-combustion engine operation to be achieved.

Compression and decompression in the cylinders means that the torque required to turn the internal-combustion engine over fluctuates considerably, as a result of which the engine speed also fluctuates considerably. Figure 2 shows a typical graph for engine speed and starter-motor current on cold start.

The starter itself must satisfy the following technical requirements:
– readiness to function at any time,
– sufficient starting power,
– long service life,

Figure 1: Starter for passenger cars and light commercial vehicles
1 DC motor,
2 Reduction gear,
3 Engaging system with solenoid switch,
4 Roller-type overrunning clutch with starter pinion.



– robust design,
– low weight and small installation space,
– maintenance-free operation.

Influencing factors
The starter must crank the internal-combustion engine at a minimum speed (starting speed) in order to achieve the thermodynamic and mechanical conditions for sustained operation. The cranking speed largely depends on the characteristics and properties of the internal-combustion engine (engine type, engine swept volume, number of cylinders, compression, bearing friction, engine oil, fuel-management system, additional loads driven by the engine), and the ambient temperature. In general, starting torque and starting rotational speed require a gradual increase in starting power as temperatures decline. However, the power supplied by a starter battery falls as temperatures drop because its internal resistance increases. This opposing relationship of electrical load requirements and available power means that the cold start constitutes an unfavorable operating scenario. This must be taken into consideration in the design of the starting system.

Classification
Passenger-car starting systems are designed for a rated power of up to 3 kW and a rated voltage of 12 V. These can start gasoline engines with up to approximately 7 liters engine displacement and diesel engines with up to approximately 4 liters.

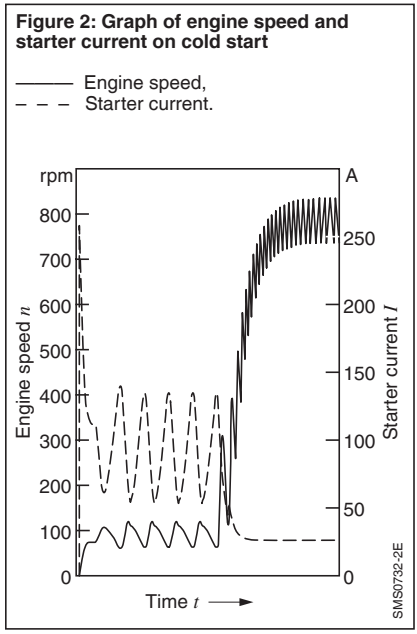
Use of lithium-ion batteries
The use of lithium-ion batteries imposes particular demands on a starter. These batteries are used because they are lower in weight, require less space for installation, and are more suitable for recuperation. Higher voltages and currents are involved compared with lead-acid batteries, meaning that the starter is subject to increased electrical and mechanical load.

Starter design and operation

A starter essentially consists of the electric motor and the engaging system with solenoid switch, engagement lever, overrunning clutch, meshing spring, and starter pinion. The engaging system ensures that the starter pinion meshes (engages temporarily) with the ring gear. In addition, the starter incorporates a reduction gear, which increases the electric-motor torque at the starter pinion.

During the start, the starter pinion engages with the ring gear by means of the solenoid switch with the assistance of the helical spline when the driving switch is operated. The starter pinion drives the internal-combustion engine via the ring gear until the engine can run at sustained operation. After a successful start the starter pinion is returned to the rest position (demeshing) by spring force and with the assistance of the helical spring while the internal-combustion engine accelerates quickly to high revs.

Figure 2: Graph of engine speed and starter current on cold start

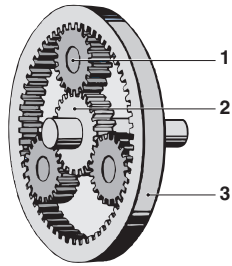


Starter motor
The starter motor is a permanent-field or electrically excited DC motor.

Reduction gear
The objective of starter development is to minimize the weight and dimensions of the starter by reducing the volume of the electric motor etc. In order to achieve the same starting power at the same time, a higher armature speed is required to compensate for the existing lower armature torque. In order nevertheless to achieve unchanged starting power, it is necessary for the now smaller armature torque of the electric motor to be compensated by a higher armature speed. The speed is matched to the engine crankshaft speed by increasing the overall step-down ratio from starter-motor armature to crankshaft with the aid of a reduction gear integrated in the starter. This reduction gear usually takes the form of a planetary gear (Figure 3). It comprises a sun gear mounted on the armature shaft, a planetary-gear carrier normally with three planet gears, and a fixed internal ring gear.

The planetary gear transmits the armature torque via the coupled planetary-gear carrier to the starter drive shaft and to the starter pinion. The high armature speed here (up to 25,000 rpm) is stepped down at a ratio of $i \approx 3$ to 5. The further reduction ratio from starter pinion to ring gear produces an engine speed of around 300 rpm.

Figure 3: Planetary gear
1 Planetary gear (planetary gears are coupled to each other via a planetary-gear carrier not pictured here),
2 Sun gear,
3 Internal ring gear.

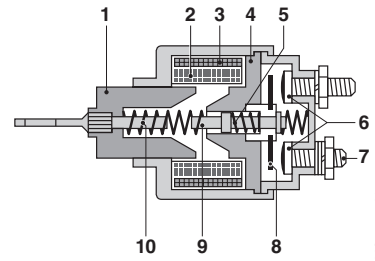


Solenoid switch
The starter current is up to 1500 A for passenger cars. Since contacts for such high currents are subjected to high loads, it is essential to use a power relay. The solenoid switch is actuated by a relatively low control current (relay current, approx. 30 A for passenger cars). For switching it is sufficient to have a mechanical switch (driving switch, start button) or a simple minirelay which is actuated by the engine ECU.

The solenoid switch used in the starter (Figure 4) consists of switch armature, pull-in winding, hold-in winding, magnetic core, contact system (contact spring, contact stud, contact bridge), switch cover with integrated connections, split switch shaft, and return spring. The solenoid switch has to fulfill two functions:
– advancing the starter pinion over the engagement lever into the ring gear and
– switching the main starter current by closing and opening the contact bridge.

The winding is divided into a pull-in winding and a hold-in winding in order to limit heating of the switch winding. Both windings are connected in parallel when the solenoid switch is activated (Figure 5). The magnetic forces of the two windings are added up. The start of the two windings are both applied to terminal 50 of the solenoid switch. The winding end of the pull-in winding is connected via the starter-motor

Figure 4: Solenoid switch
1 Switch armature, 2 Pull-in winding,
3 Hold-in winding, 4 Magnetic core,
5 Contact spring, 6 Contacts,
7 Electrical connection, 8 Contact bridge,
9 Armature shaft (divided), 10 Return spring.



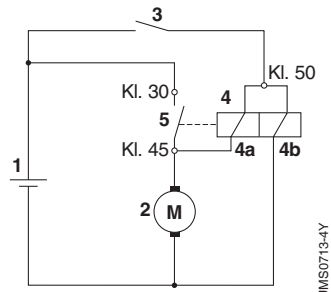
armature, that of the hold-in winding is connected directly to the ground potential. When the battery positive terminal is connected to terminal 50 of the solenoid switch (driving switch in start position), the switch armature is pulled axially into the housing by the magnetic force generated by the pull-in winding and hold-in winding. This movement pushes the starter pinion forwards via the engagement lever towards the ring gear (Figure 6). Only once the switch armature is almost completely pulled in is the contact bridge closed and the main starter current switched on. This prevents the starter motor beginning to rotate before the starter pinion has engaged in the ring gear.

As the two winding ends for the pull-in winding are now connected to positive, current flows only in the hold-in winding. The lower magnetic force of the hold-in winding is sufficient to hold the switch armature securely until the driving switch is opened.

The pull-in and hold-in windings must have the same number of turns. Otherwise the solenoid switch may lock during the deactivation process because the two windings now connected in series are supplied in reverse via terminal 45. The same numbers of turns ensure that the magnetic fields of the coils now passed through in opposite directions compensate each other and the solenoid switch switches off reliably.

Figure 5: Solenoid-switch circuit

1 Battery,
2 Starter motor,
3 Driving switch or switch of a minirelay for automatic starting system,
4 Solenoid switch,
4a Pull-in winding, 4b Hold-in winding,
5 Contact bridge.

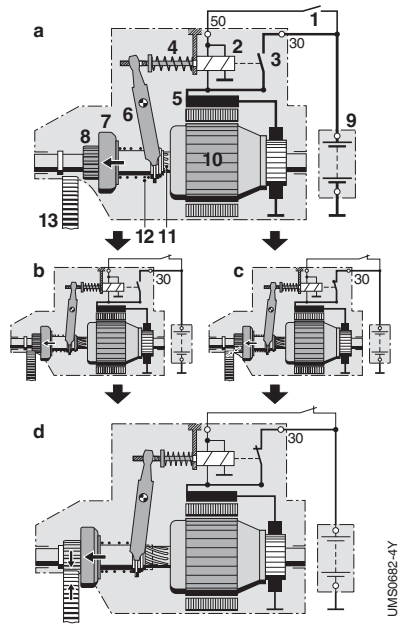


Meshing of the pre-engaged-drive starter

The pre-engaged-drive starter has become a worldwide standard for the passenger car, as reliable operation over the entire operating range is guaranteed with this engagement principle. On pre-engaged-drive starters, the engagement travel is made up of a lever-travel phase and a helical-travel phase. When the driv-

Figure 6: Working phases of a pre-engaged-drive starter

a) Rest position,
b) Tooth on gap,
c) Tooth on tooth,
d) Engine is cranked.
1 Ignition/starting switch (driving switch),
2 Solenoid switch, 3 Contact bridge,
4 Return spring,
5 Excitation winding (series winding),
6 Engagement lever,
7 Roller-type overrunning clutch,
8 Starter pinion, 9 Battery, 10 Armature,
11 Helical spline, 12 Meshing spring,
13 Engine-flywheel ring gear.



ing switch is closed terminal 50 of the solenoid switch is connected to the battery positive terminal. The solenoid-switch armature attracts the engagement lever, which in turn pushes the overrunning clutch with the starter pinion via the helical spline against the internal-combustion engine's ring gear (lever travel). The starter-motor armature still does not rotate in this phase as the main current for the armature and excitation windings is not yet switched on (Figure 6a).

If a pinion tooth meshes directly with a tooth gap on the ring gear (tooth-gap positioning) when the starter pinion meets the ring gear, the starter pinion engages as far as the movement of the solenoid switch allows (Figure 6b). If the starter-pinion tooth strikes a tooth on the ring gear (tooth-tooth positioning, Figure 6c) when the starter pinion meets the ring gear – this occurs approximately 80 % of the time – the switch armature tensions the meshing spring via the engagement lever, as the starter pinion can travel no further axially. Once the pinion travel limit generated by the solenoid switch has been reached, the contact bridge of the switch armature closes and the starter-motor armature begins to rotate. In the case of tooth-gap positioning, the rotating starter motor screws the starter pinion fully into the ring gear via the helical spline (helical travel). Starting from a tooth-tooth position, the starter motor turns the starter pinion in contact with the ring gear until a pinion tooth finds a tooth gap in the ring gear and the pretensioned meshing spring pushes the starter pinion and overrunning clutch forward.

The sloping teeth provided at the front of the starter pinion permits an earlier start of the axial engagement movement and consequently produces a greater alignment of pinion tooth and ring-gear tooth at the moment of the first power transmission. This reduces wear.

When the driving switch is opened, the return spring pushes the switch armature – and the starter pinion and overrunning clutch via the engagement lever – back into the rest position. The overrunning torque caused by the friction of the overrunning clutch generates an axial force in conjunction with the helical spline, which supports the starter-pinion's demeshing process.

Overrunning clutch

In all starter designs, the drive torque is transmitted by an overrunning clutch (overriding clutch). It is located between the starter motor and the starter pinion and has the following functions:

– Transmitting the torque from the starter motor to the pinion as long as the starter motor cranks the internal-combustion engine.
– Disengaging the connection between starter pinion and drive shaft as soon as the internal-combustion engine by its own efforts accelerates to higher revs than the starter motor.

Roller-type overrunning clutch

Overrunning clutches for starters are usually frictional in design (roller-type overrunning clutch, Figure 7). The assembly essentially consists of

– the driver with the roller race,
– the cylindrical rollers,
– the springs, and
– the starter pinion with the inner race.

The driver is connected via the helical spline to the drive shaft. The cylindrical rollers can move on the roller race and establish the frictional connection between the inner race of the starter pinion and the outer-rotating clutch shell of the driver. The springs press the cylindrical rollers into the narrowing part between the driver's roller race and the inner race. When the starter motor starts, the inner race is also move by the frictional forces acting on the cylindrical rollers. The greater the torque, the further the cylindrical rollers are moved into the narrowing part of the roller race, and as a result the transmissible frictional forces increase proportionally to the torque.

When the overrunning operation is initiated, the cylindrical rollers are moved into the widening part of the roller race, whereby with relatively small force they are in contact with the slipping inner race and the roller race. The overrunning torque that develops is dependent on the spring force and relatively small compared with the torques acting in the drive

direction. However, it does impact on the idle speed of the starter-motor armature that is obtained.

Usually the overrunning clutch is moved together with the pinion during engagement axially in the direction of the ring gear. However, there are starter versions in which only the pinion is moved and the overrunning clutch is axially fixed ("stationary overrunning clutch").

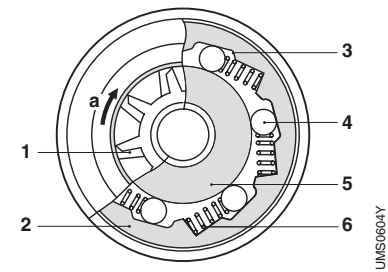
Demeshing
When the driving switch is released, the voltage at the solenoid-switch windings returns to zero. The magnetic force abates and the return spring presses the switch armature back, as a result of which initially the primary-current contact (contact bridge) opens and the starter current is interrupted. The return spring completes the retraction of the starter pinion, disengaging it from the ring gear. The demeshing process is supported by the axial forces acting in the helical spline. The starter freewheels to a standstill and the entire engagement mechanism returns to the rest position.

Switch-off function
The switch armature is connected to the engagement lever in such a way that there is a certain amount of free travel or "play". If the internal-combustion engine fails to achieve sustained operation, the start must be aborted. At this point, the starter pinion and the ring gear are under full load and the starter pinion is fully engaged. When the switch current is switched off, there must be sufficient switch-armature free travel to allow it to open the primary-current contacts. If this is not the case, the engagement lever holds the switch armature in place. The primary-current contact remains closed and it is not be possible to abort the start.

Configuration of a starter

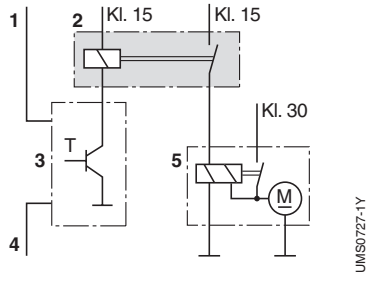
The most important boundary conditions which must be taken into consideration in the configuration of a starter are:
- The minimum starting temperature, i.e., the lowest engine temperature at which a start must still be possible.
- The resistance to rotation of the engine, i.e., the required torque at the crankshaft including all supplementary loads.
- The requirement minimum engine speed at the minimum starting temperature.
- The transmission ration between the starter and the crankshaft.
- The nominal voltage of the starting system.
- The properties of the starter battery.
- The resistance of the supply leads between the battery and the starter and the contact resistances of terminals and switching elements (isolating switches, etc.).
- The starter's speed/torque characteristic.
- The maximum permissible voltage dip in the vehicle electrical system at which the operating capability of the vehicle electronics is still guaranteed.

Figure 7: Roller-type overrunning clutch
1 Starter pinion,
2 Clutch shell,
3 Driver with roller race,
4 Cylindrical roller,
5 Pinion shaft (inner race),
6 Spring.
a Direction of rotation.



In view of these boundary conditions, the starter cannot be considered in isolation. As an integral part of the overall system of internal-combustion engine including its ancillaries, the vehicle electrical system with battery and cabling as well as the starter itself, the starter must be matched to the other components.

Figure 8: Automatic starting system (circuit diagram)
1 Starting signal from driver,
2 Ballast relay,
3 ECU,
4 Park-Neutral position or clutch signal,
5 Starter.



Triggering the starter

Conventional control
During conventional starts, the driver connects the battery voltage (driving switch in starting position) to the starter solenoid switch. The switch current generates in the solenoid switch the switch force which on the one hand moves the starter pinion towards the ring gear and on the other hand switches on the starter's primary current. The starter is switched off by means of the driving switch, which opens and disconnects the voltage from the solenoid switch.

Automatic starting systems
The high demands on vehicles with regard to convenience, safety, quality, and low noise levels have resulted in an increase in the use of automatic starting systems. An automatic starting system differs from a conventional one by virtue of additional components: one or more ballast relays, as well as hardware and software components (e.g., an ECU) to control the starting sequence.

The driver then no longer directly controls the starter relays current; instead the ignition key is used to send a request signal to the control unit, which then performs a series of checks before initiating the starting sequence. The safety check comprises a wide range of options and can verify any of the following:
– Is the driver authorized to start the vehicle (theft-deterrence feature)?
– Is the internal-combustion engine stationary (prevents the pinion from meshing with the moving ring gear)?
– Is the state of charge (in relation to engine temperature) sufficient to carry out the start?
– For automatic transmissions, is the selector lever in the Park position or, for on manually shifted transmissions, is the clutch disengaged?

When the check has been completed successfully, the control unit initiates the start. During the start the starting system compares the engine speed with a desired speed for the internal-combustion engine. Once the engine reaches the desired speed, the ECU switches the starter off. This achieves the shortest possible starting time, reduces the noise levels, and lessens starter wear.

Start/stop system

An automatic starting system can also be used as the basis to implement start/stop operation. Here, the internal-combustion engine is switched off when the vehicle is not in motion taking into account defined temperature inputs and the state of charge of the battery, e.g., at traffic lights on red, and restarts automatically when required.

The operation of a start/stop system – consisting of start/stop starter motor, ECU with start/stop function, electronic battery sensor, deep-cycle-resistant battery, pedals with sensors – presupposes a higher-level activation by means of which the switch-off and restart strategy is implemented.

A start/stop system requires an electrical energy management system that incorporates battery-charge detection. Measures may also be required to stabilize the vehicle electrical system during the starting phase to avoid unacceptable voltage drops. The control equipment and starting system must therefore be matched. The control units have to fulfill their function at a significantly lowered supply voltage.

Demands imposed on start/stop starters

Driven by the efforts of the automotive industry to reduce CO_2 emissions, the global market share of vehicles with start/stop functionality has risen. This is accompanied by demands imposed on starting comfort (start duration and starting noise), on starter service life, and on availability. One technical challenge with regard to start/stop starters is to optimize the rapid-starting performance of the respective internal-combustion engine (repeatable rapid starting). To be able to satisfy the demands with regard to service life and more rapid starting to reduce noise generation, the deep-cycle resistance of start/stop starters is increased significantly compared with conventional starters. Such starters exhibit features like optimized pinion geometry and pinion resilience, load damping, and long-life carbon-brush systems. Starters for start/stop operation achieve in conjunction with a high-quality ring gear more than 330,000 starting cycles.

If in a start/stop system the engine stop is immediately followed by a start request (e.g., because the traffic lights turn green at the very moment of the engine stop) this is referred to as “Change of Mind” (CoM). In such a situation the length of time until the engine restart is reduced whereby with the restart the system does not wait for the engine to be at a standstill but instead a start occurs while the engine is still moving. This results in further demands imposed on the starter, particularly with regard to its capacity to mesh with a moving ring gear. Above all meshing with a return-swinging ring gear generates high loads in the drivetrain and wear at the interface between starter pinion and ring gear. In a positive CoM situation (engine turns in the drive direction) additional component load and wear is also accompanied by increased noise emissions. Knowledge of the rundown behavior of the internal-combustion engine and with it the loads to be expected during CoM starts is of vital importance when it comes to designing CoM-capable start/stop systems.

Starters for heavy commercial vehicles

Application

Starters for heavy commercial vehicles are used in trucks as well as in construction machinery, agricultural machinery, locomotives, and stationary engines. These starters have a similar basic design to passenger-car starters (see Starters for passenger cars and light commercial vehicles), sharing among others the following common components (Figure 1):
– DC motor,
– reduction gear,
– engaging system with solenoid switch,
– roller-type overrunning clutch and starter pinion.

Commercially-vehicle starters are typically based on four-pole electrically excited direct-current machines. Starters for large engines (HEF and TB/TF sizes) are designed as six-pole electrically excited machines.

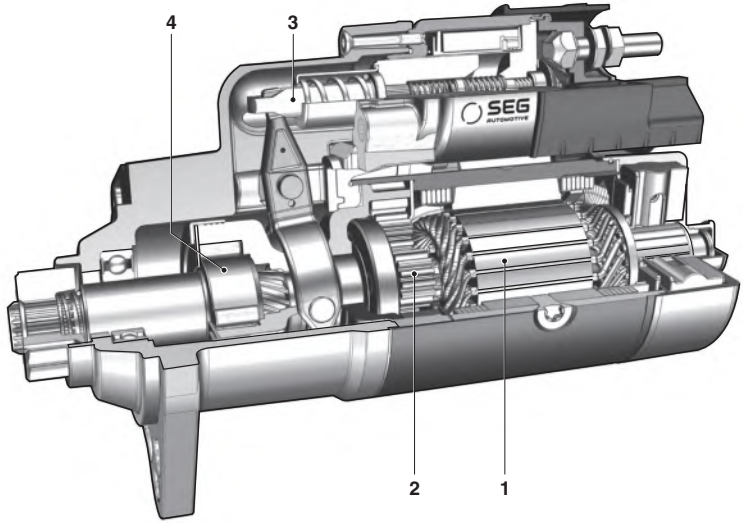
Power requirements

Typical commercial vehicles are equipped with diesel or gasoline engines (sometimes even gaseous-fuel engines) with a displacement of approx. 3 l through 16 l, and for special applications even up to approx. 150 l. The starter types developed for these requirements cover a power range (specified maximum power output) of approx. 2.3 kW through approx. 10 kW in 12 V and 24 V vehicle electrical systems. Furthermore, even starters with up to 25 kW and 110 V are available for special applications.

For large engines up to 150 l displacement HEF and TB/TF size starters (no longer used for new applications) can also be used in parallel starting systems, in which up to three starters supply their power to a common ring gear (Figure 2).

Figures 3 and 4 provide an overview of the starter types used in the commercial-vehicle sector.

Figure 1: Starter for heavy utility vehicles with four-pole electrically excited direct-current machines
1 DC motor, 2 Reduction gear, 3 Engaging system with solenoid switch, 4 Roller-type overrunning clutch and starter pinion.



Operating principle

Torque is transmitted via roller-type overrunning clutches which, depending on the starter size, are moved along with the pinion or are designed to be stationary. To satisfy special requirements, commercial-vehicle starters have special features which are described in the following.

Two-stage meshing

The starter pinion meshes with the ring gear prior to the actual starting operation. The starter torque can be transmitted only when the pinion has become sufficiently aligned with the ring gear.

To guarantee this function with minimal wear even with high starter torque over the entire service life, commercial-vehicle starters are equipped from a rated power of approx. 5 kW with a two-stage meshing system. Here, after the switching relay has been activated (Figure 5), the starter pinion is already turned slowly during the meshing operation to be able to mesh with a gap before the primary current is switched, and then the full starter torque is available. The torque of the first meshing stage is for this purpose reduced by a winding in the solenoid switch (pull-in winding), which is connected as a series resistor to the electric motor (Figure 5a). On successful meshing the switch shaft switches off the pull-in winding shortly before the end point is reached via the NC contact. Shortly after the primary current is switched on via the main contact (Figure 5b).

In some starter types the synchronization between pinion movement and main

Figure 2: Parallel starting system

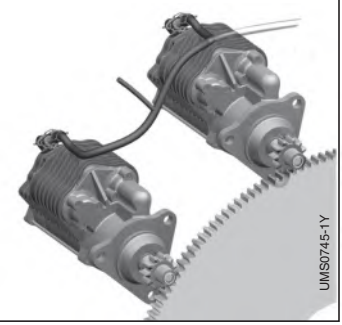


Figure 4: Starter types for commercial vehicles (12 V)

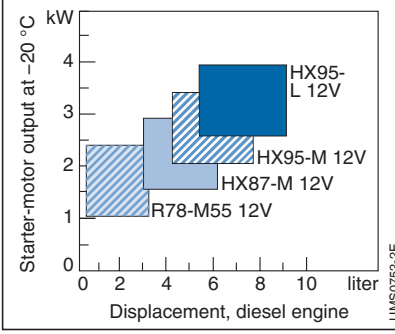
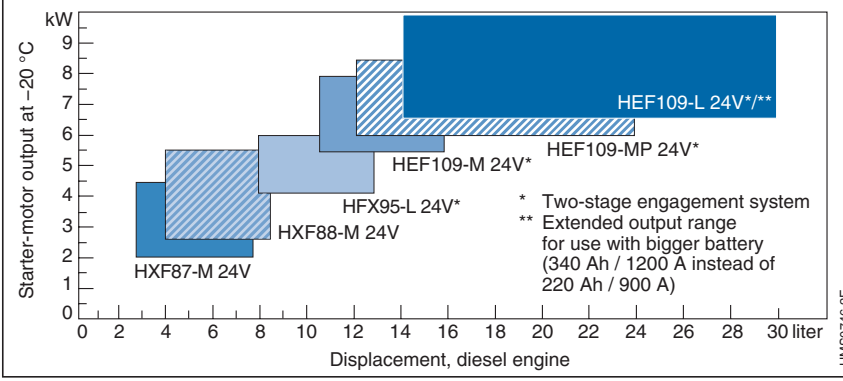


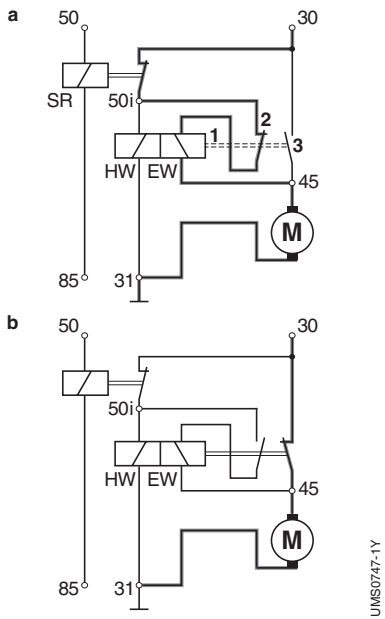
Figure 3: Starter types for commercial vehicles (24 V)



contacts is configured in such a way that the required meshing depth is not only dynamically achieved but also geometrically predefined. These starter types perform in the event of a “corner to corner” connection between pinion and ring gear a “blind connection”, in which the second stage is not switched. The starting operation must then be interrupted by the operator (or the ECU) and repeated.

Figure 5: Current flow for two-stage meshing system

a) First stage: series resistor (EW) limits the armature current → slow turning of pinion for meshing.
b) High armature current after closing of main contacts → full starting torque to crank the engine.
1 Switch shaft,
2 NC contact,
3 Main contact.
SR Switching relay,
HW Hold-in winding,
EW Pull-in winding,
30, 31, 45, 50, 50i, 85 Terminal designation.



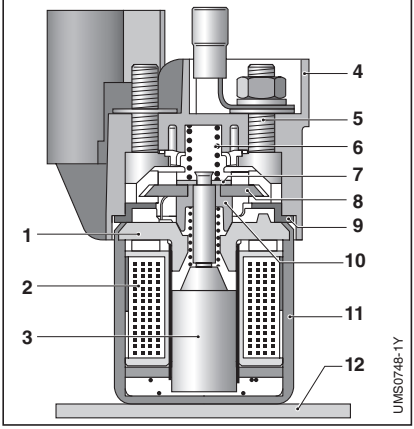
Switching relay

The relay activation currents required for starter activation are up to 60 A for one-stage meshing systems and up to 200 A for two-stage meshing systems. To be able to switch these currents, an additional switching relay (“IMR”, integrated mechanical relay, Figure 6), which is actuated with an activation current of approx. 2 A, is used in all two-stage systems. This relay is mounted on the starter (see Figure 7). It is a compact power relay consisting of a magnetic circuit with simple winding and a contact system with angled contact plate and two contact screws. For contact making the armature draws into the excited magnetic circuit and with the contact plate completes the circuit between the two contact screws. After the excitation current has been switched off, the return spring lifts the contact plate back off the contact screws and pushes the armature back to the rest position.

The switching relay is optionally also available for one-stage systems to reduce the activation current at the system end.

Figure 6: Switching relay (IMR)

1 Magnetic core, 2 Winding, 3 Armature,
4 Switch cap, 5 Contact screw,
6 Return spring, 7 Insulating shim,
8 Contact plate, 9 Seal,
10 Insulating bushing, 11 Housing,
12 Flange plate.



Special requirements for commercial-vehicle starters

Ambient conditions
The engine compartments of commercial vehicles are subjected to heavy dirt, dust and water bombardment both in agricultural- and construction-machinery applications and in on-highway applications. To ensure robust operation under these boundary conditions, commercial-vehicle starters are increasingly being equipped with a free-ejecting meshing drive. The starter is sealed here by a shaft seal between pinion shaft and starter housing against the engine compartment.

Free-ejecting commercial-vehicle starters can optionally also be completely sealed against water ingress; they are then suitable for use in forded conditions or with a flooded engine compartment. A pressure-equalization element is used here to reduce air-pressure fluctuations inside the starter caused by ambient and temperature conditions.

Protection against misuse
Because of the varied conditions of use and the prerequisite of maximum availability, some applications necessitate a feature that provides protection against misuse, preventing the starter from incurring damage due to excessive activation.

For this purpose, starters are available for mid-size and heavy diesel engines with the option of a built-in thermostatic switch (Figure 7) in the electric motor which interrupts an excessively long cranking operation before the starter can incur damage due to overheating.

This safety feature is self-resetting and reversible, i.e., a new starting operation can be initiated after the starter has cooled down.

In TB/TF Series starters thermostatic switches are optionally available in the carbon brushes and in the relay winding as well to provide further safeguarding in other operating conditions.

Potential-free housing
In commercial vehicles and also in passenger cars the engine block or the vehicle chassis is usually used as frame potential for the electrical connection between the battery and the starter. However, some special applications (e.g., in ships, locomotives, and special vehicles with readily flammable cargos) require the vehicle chassis to be isolated (potential-free). For this purpose commercial-vehicle starters are optionally equipped with an “isolated return line” (Figure 8), in which all the connections are insulated against the starter housing.

The negative-side connection to the vehicle electrical system is then established via a special terminal 31 with a separate supply line to the negative battery terminal.

Figure 7: Starter with thermostatic switch
1 Switching relay, 2 Solenoid switch,
3 Thermostatic switch.

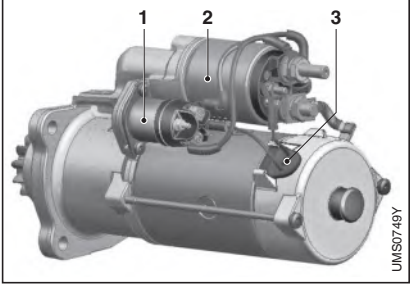
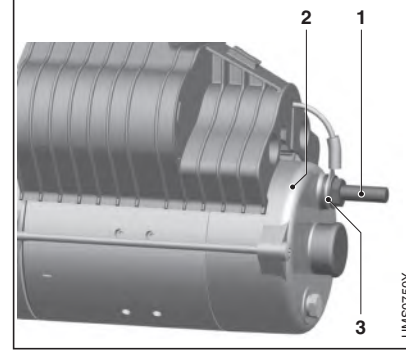


Figure 8: Isolated return line
1 Terminal 31, 2 Commutator bearing,
3 Insulating part.



Wiring harnesses and plug-in connections

Wiring harnesses

Requirements
The purpose of the wiring harness is to distribute power and signals within a motor vehicle. A wiring harness in the present-day, mid-class passenger car with average equipment has approximately 750 different lines, their length totaling around 1500 meters (Table 1). In recent years, the number of contact points has practically doubled due to the continuous rise in functions in the motor vehicle. A distinction is made between engine-compartment wiring harnesses and between body, front-end, rear-end, cockpit, roof, door, and tank wiring harnesses (Figure 1, Table 2). Above all, engine-compartment wiring harnesses are subject to high temperature, vibration, and media loads.

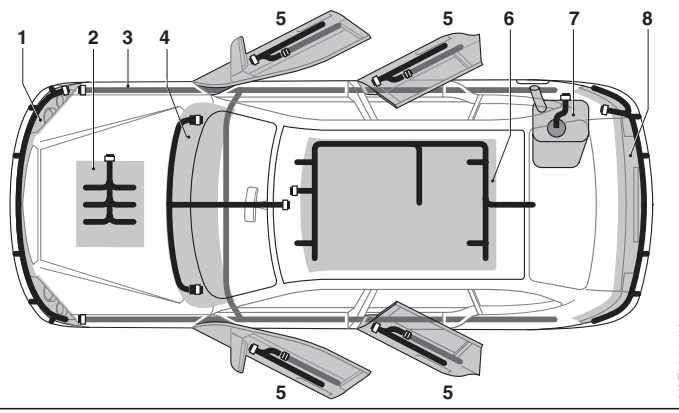
Wiring harnesses have considerable influence on the costs and quality of a vehicle. The following points must be taken into consideration in wiring-harness development:
– imperviousness to dust and media,
– electromagnetic compatibility,
– temperatures,
– protection of the wires against damage,
– wire routing,
– ventilation of the wiring harness.

It is therefore necessary to involve wiring harness experts as early as in the system definition stage. Figure 2 shows a wiring harness that was developed as a special intake-module wiring harness. Thanks to the optimization of routing and security in conjunction with engine and wiring-harness development, it was possible to achieve an advancement of quality as well as to yield cost and weight advantages.

Dimensioning and selection of materials
The most important tasks for the wiring harness developer are:
– dimensioning of line cross-sections,
– material selection,
– selection of suitable plug-in connections,
– routing of lines under consideration of ambient temperature, engine vibrations, acceleration and EMC,
– consideration of the environment in which the wiring harness is routed (topology, assembly stages in vehicle manufacture and equipment on the assembly line).

Line cross-sections
Line cross-sections are defined based on permissible voltage drops. The lower

Figure 1: Wiring-harness modules
1 Front end,
2 Engine compartment,
3 Body,
4 Cockpit,
5 Doors,
6 Roof,
7 Tank,
8 Rear end.



cross-section limit is determined by the line strength. Convention has it that no lines with a cross-section of less than 0.5 mm^2 are used. With additional measures (e.g. supports, protective tubes, tension relief), even a cross-section of 0.35 mm^2 may be permissible.

Table 1: Complexity of wiring harnesses (typical values for complete wiring harness)

	Small vehicle	Medium class	Luxury class
Number of plug connectors	70	120	250
Number of contact points	700	1500	3000
Number of wires	350	750	1500
Total length of wires in meters	700	1500	3200

Table 2: Data for wiring harnesses (typical values for vehicle with average equipment)

	Body wiring harness	Cockpit wiring harness
– Different modules	54	23
– Wires	528	57
– Twisted wires	24	24
– Total length (in meters)	1370	226
– Different contact parts	72	17
– Assorted components (e.g. plug housings, cable ducts, coverings)	227	63

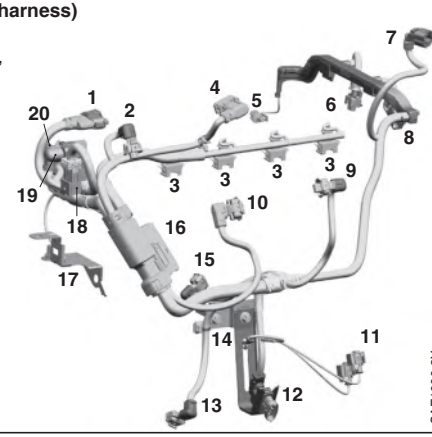
Materials

Copper is usually used as the conductive material on account of its high conductivity. Recent trends have involved the increased use of aluminum on account of its weight and cost advantages above all for line cross-sections greater than 2.5 mm^2. The insulation materials for the lines are defined by the temperature to which they are exposed. It is necessary to use materials that are suitable for the high temperatures of continuous operation. Here, the ambient temperature must be taken into consideration as much as the heating caused by the flow of current. The materials used are thermoplastics (e.g. PE, PA, PVC), fluoropolymers (e.g. ETFE, FEP), and elastomers (e.g. CSM, SIR).

If the lines are not routed past particularly hot parts (e.g. exhaust pipe, exhaust-gas recirculation) in the engine topology, one of the criteria for the selection of the insulation material and the cable cross-section could be the derating curve of the contact with its associated line. The derating curve represents the relationship between current, the temperature increase that it causes, and the ambient temperature of the plug-in connection. It can be taken from the relevant technical customer documentation for the contact. Normally, the heat generated in the contacts can only be carried away along the lines themselves. What is also to be noted is that at high usage tempera-

Figure 2: Intake-module wiring harness (example of an engine-compartment wiring harness)

Connector for
1 Ignition-coil module, 2 Channel deactivation,
3 Fuel injectors, 4 Throttle device,
5 Oil-pressure switch,
6 Engine-temperature sensor,
7 Intake-air temperature sensor,
8 Camshaft sensor,
9 Canister-purge valve,
10 Intake-manifold pressure sensor,
11 Charge-current indicator lamp,
12 λ sensor downstream of catalytic converter,
13 Speed sensor,
14 Terminal 50, starter switch,
15 Knock sensor, 16 Engine ECU,
17 Engine ground,
18 Separable connector for engine and transmission wiring harnesses,
19 λ sensor upstream of catalytic converter,
20 Exhaust-gas recirculation valve.



tures (>100 °C) the contact materials can give way to the applied mechanical stresses (metal relaxation). It is possible to influence the relationships described by means of larger line cross-sections and the use of suitable contact types and more noble surfaces (e.g. gold, silver) and thus higher limit temperatures. For highly fluctuating current intensities, it is often useful to measure the contact temperature on an application-specific basis.

Line routing and EMC measures
Lines should be routed in such a way as to prevent damage and line breaks. This is achieved by means of fasteners and supports for the wiring-harness branches. Vibration loads on contacts and plug-in connections are reduced by fastening the wiring harness as close to the plug as possible and at the same level as the vibration where possible. The line routing must be determined in close cooperation with the engine or the vehicle developer.

To avoid EMC problems, it is recommended to route sensitive lines and lines with steep current flanks separately, making maximum use of the available space. Shielded lines are not straightforward to produce and are therefore expensive. They also need to be grounded. The twisting in pairs of signal lines is a more cost-favorable and effective measure.

Line protection
Lines need to be protected against chafing and against making contact with sharp edges and hot surfaces. Adhesive tapes are used for this purpose. The level of protection is determined by the interval and winding density. Corrugated tubing (material savings from corrugation) with the necessary connecting pieces is often used as line protection. However, tape fixing is still an essential means of preventing movement of individual lines inside the corrugated tube. Optimal protection is offered by cable ducts.

Wiring harnesses are at risk from rodents. A remedy may be provided by rodent-resistant, extruded plastic tubing.

Plug-in connections

Function and requirements
Electrical plug-in connections must establish a reliable connection between different system components and thereby ensure the safe functioning of the systems under all operating conditions. They must be designed in such a way as to withstand the multitude of loads to which they are exposed over the entire service life of the motor vehicle. Examples of such loads are
– vibration loads,
– temperature fluctuations,
– high and low temperatures,
– humidity and moisture loads and splash water,
– aggressive liquids and corrosive gases, and
– micro-movement of the contacts with resulting fretting corrosion.

These loads can increase the contact resistances of the contacts up to the point of total interruption.

Even the insulation resistances can decrease and result in the short-circuiting of neighboring lines. Electrical plug-in connections must therefore exhibit the following properties:
– low contact resistances of the current-carrying parts,
– high insulation resistance between the current-carrying parts of different voltage potential,
– high imperviousness to water, moisture, and salt spray.

Aside from the physical properties, plug-in connections must satisfy further requirements for the area of application, such as
– easy, error-free handling in automotive assembly,
– safe reverse-polarity protection,
– secure and noticeable locking and easy unlocking,
– robustness and compatibility with automatic machine handling in wiring-harness production and transportation.

The high integration density of electronics in the motor vehicle places high demands on a car's plug-in connections. Not only do they carry high currents (e.g. activation of ignition coils), they also carry analog

signal currents with low voltage and current intensity (e.g. signal voltage of the engine-temperature sensor). Throughout the service life of the vehicle, the plug-in connections must ensure the reliable transmission of signals between ECUs and to the sensors and actuators.

The increasing demands of emission-control legislation and active vehicle safety are forcing the ever more precise transmission of signals through the contacts of the plug-in connections. Many parameters must be taken into account in the conception, design, and testing of a plug-in connection (e.g. environmental influences, imperviousness, contact forces, contact form, surface quality).

The most common cause of failure of a plug-in connection is wearing of the contact caused by vibrations or temperature change. The wear promotes oxidation. This results in an increase in ohmic resistance - the contact may, for example, be subjected to thermal overload. The contact part may be heated beyond the melting point of the copper alloy. In the case of highly resistant signal contacts, the vehicle controller often detects a plausibility error by comparison with other signals; the controller then enters fault mode. These problem areas in the plug-in connection are diagnosed by the on-board diagnostics (OBD) required by emission-control legislation. However, it is difficult to diagnose the error in the service workshops because this defect is displayed as being a component failure. It is only possible to diagnose the faulty contact indirectly.

For the assembly of the plug-in connection, there are various functional elements on the plug housing intended to ensure that the cables with their crimped contacts can be joined to the plug-in connection reliably and defect-free. Modern plug-in connections have a joining force of less than 100 N so that the assembly operative is able to reliably join the plug to the component or ECU interface. The risk of plug-in connections being connected to the interface incorrectly increases with higher connecting forces. The plug would come loose during vehicle operation.

Table 3: Uses of plug-in connections

	No. of pins	Special features	Application
Low-pin-count	1 to 10	No joining-force support	Sensors and actuators (many different requirements)
High-pin-count	10 to 300	Joining-force support by slide, lever, module construction	ECUs (several, similar requirements)
Special plugs	any	e.g. integrated electronics	Special applications (individual, matched requirements)

Design and types

Plug-in connections have different areas of application (Table 3). These are characterized by the number of pins and ambient conditions. There are four different classes of plug-in connection: direct engine attachment, attachment close to the engine, transmission attachment, and body attachment. Another difference is the temperature class of the installation location.

High-pin-count plug-in connections

High-pin-count plug-in connections are used for all ECUs in the vehicle. They differ in their number of pins and the geometry of the pins (Table 4). Figure 3 shows a typical design of a high-pin-count plug-in connection. An example of the mating component of the plug is pictured in Figure 4. The high-pin-count male connectors are usually soldered in or stitched in on the ECU board.

The complete plug-in connection is sealed against the plug connector of the associated ECU by a circumferential radial seal in the plug housing. This, together with three sealing lips, ensures a reliable seal against the ECU sealing collar.

The contacts are protected against the ingress of humidity along the cable by a flat seal, through which the contacts are inserted and the line crimped to them. A silica-gel mat or silica mat is used for this

purpose. Larger contacts and lines may also be sealed using a single-core seal (see Low-pin-count plug-in connections).

When the plug is assembled, the contact with the line attached is inserted through the flat seal that is already in the plug. The contact slides home into its position in the contact carrier. The contact latches on its own by a locking spring that engages in an undercut in the plastic housing of the plug. Once all contacts are in their final position, a slide pin is inserted to provide a second contact safeguard, or secondary lock. This is an additional security measure and increases the retaining force of the contact in the plug-in connection. In addition, the sliding movement is a means of checking that the contacts are in the correct position.

The operating force of the plug-in connection is reduced by a lever and a slider mechanism. For this purpose the studs provided on the side of the plug shroud in Figure 4 are engaged by the slider links (Figure 3) when the plug is connected to the interface. When the lever mounted in the plug housing is repositioned, the slide is pushed via toothing transversely to the plug-in direction into the plug housing and uses the links on both sides to draw the plug into the interface by the studs.

Low-pin-count plug-in connections

Low-pin-count plug-in connections (Figure 5) are used for actuators (e.g. fuel injectors) and sensors. Their design is similar in principle to that of a high-pin-count plug-in connection. The operating

Table 4: Flat connectors (examples)

		Pin thickness in mm		
		0.4	0.6	0.8
Pin width in mm	0.5	x		
	0.6		x	
	1.2		x	
	1.5		x	
	2.8			x
	4.8			x

Figure 4: Component side of plug-in connection

1 Plug shroud, 2 Stud.

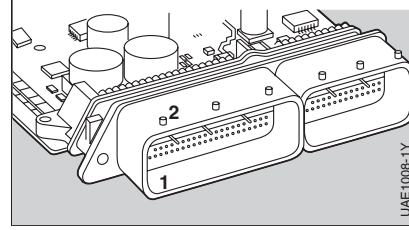
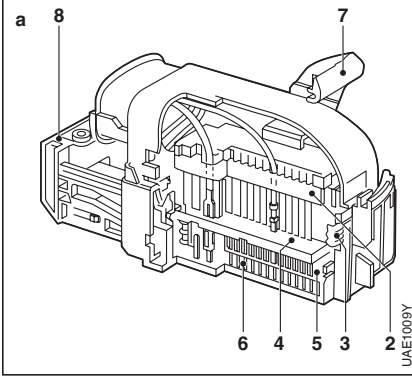
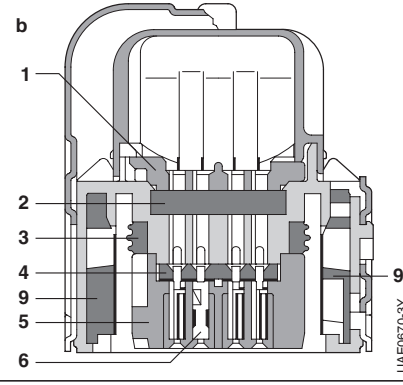


Figure 3: High-pin-count plug-in connection

a) View, b) Cross-section.
1 Pressure plate, 2 Flat seal, 3 Radial seal, 4 Slide pin (secondary lock), 5 Contact carrier, 6 Contact, 7 Lever, 8 Slider mechanism, 9 Slider links.





force of the plug-in connection is not usually supported.

The connection between a low-pin-count plug contact system and the interface is sealed with a radial seal. Inside the plastic housing, however, the lines are sealed with single-core seals secured to the contact.

Contact systems

The electrical contact system inside a plug-in connection is formed as a rule by a jack (on the plug side) and a contact pin, or a flat blade, or a pin (on the component side).

Figure 5: Low-pin-count plug-in connection

1 Contact carrier, 2 Housing,
3 Radial seal, 4 Interface, 5 Flat blade.

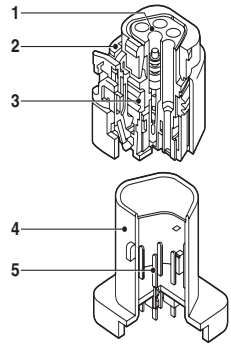
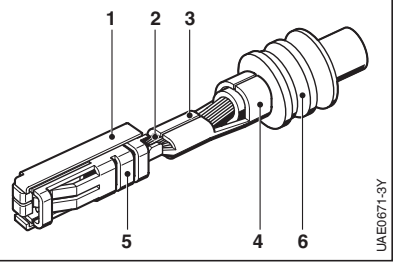


Figure 6: Contact

1 Steel overspring, 2 Single core,
3 Conductor crimp, 4 Insulation crimp,
5 Meander, 6 Single-core seal.



There are different type series for the various fields of use for Bosch plug-in connections. Depending on the conditions of use, the specially suited contacts are used in these. Figure 6 shows by way of example a two-part contact which is used in motor vehicles in components which are subject to particular vibration loads.

The inner part – the live part – is pressed from a high-quality copper alloy. It is protected by a steel overspring, which at the same time increases the contact forces of the contact by means of an inwardly acting spring element. A catch arm pressed out from the steel overspring engages the contact in the plastic housing part.

Contacts are coated with tin, silver or gold, depending on requirements. To improve the wear characteristics of the contact point, not only are different contact coatings used but also different structural shapes. Different decoupling mechanisms are integrated into the contact part to decouple cable vibrations from the contact point (e.g. meandering routing of the supply lead).

The cables are crimped onto the contact. The crimp geometry on the contact must be adapted to the cable concerned. Pliers or fully automatic, process-monitored crimping presses with contact-specific tools are available for the crimping process.

Electromagnetic compatibility

Requirements

Electromagnetic compatibility (EMC) is the general term for the target state in which electrotechnical devices and state-of-the-art technologies (e.g. wireless systems) do not have unintended mutual interference from electrical, magnetic or electromagnetic effects.

Particularly in vehicle electronics, the importance of electromagnetic compatibility is increasing due to the higher amount of electrical equipment and the use of new technologies, such as electric and hybrid drives with new energy accumulators (e.g. high-performance batteries and fuel cells) on the one hand and mobile communication systems (e.g. phone, navigation and Internet) on the other. This leads to a rising level of complexity in today's automobiles (Figure 1).

Electronic power train systems installed in the vehicle (e.g. engine and transmission control, electric drive), safety systems (e.g. antilock braking system, driving-dynamics control system, and airbag), comfort and convenience electronics (e.g. climate control, electrical adjusting devices) and mobile communication systems (e.g. radio, navigation, Internet) are arranged next to each other in close proximity. The associated high density and number of fast-switching, high-performance electronic components on the one hand and the requirements of today's communication on the other pose a great and, in part, new challenge to electromagnetic compatibility (Figure 2).

Figure 2: Electromagnetic compatibility in motor vehicles

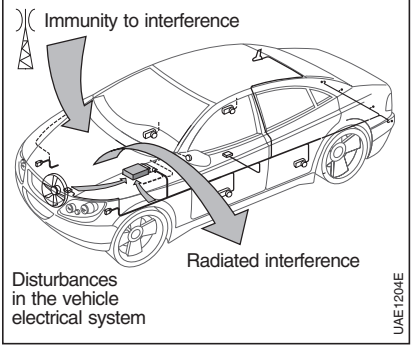
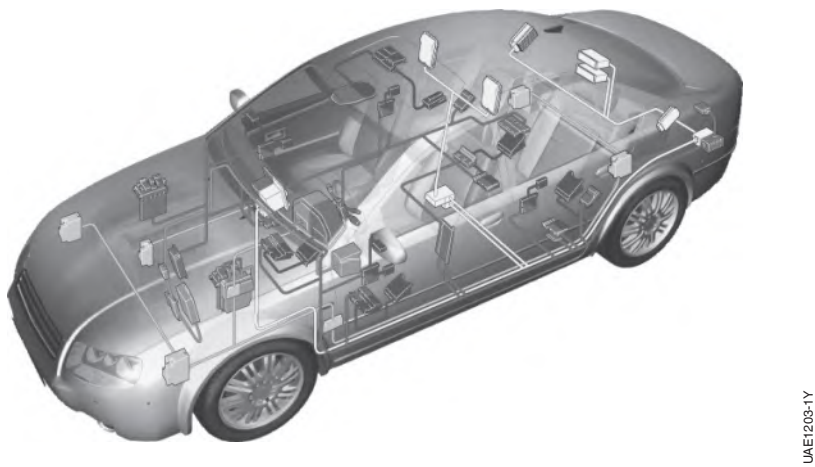


Figure 1: Electronic systems and wiring harness in a modern motor vehicle



Interference emission and interference immunity

Sources of interference in the DC vehicle electrical system

Ripple in vehicle electrical system

In a conventional vehicle with combustion engine, the alternator supplies the vehicle electrical system with rectified three-phase current. Although the current is smoothed by the vehicle battery, a residual ripple remains. In addition, the energy requirements of the electrical and electronic systems influence the DC voltage supply.

The amplitude of the ripple in the vehicle electrical system depends on the load on the vehicle electrical system and the wiring. Its frequency changes according to alternator speed and the consumer behavior. The fundamental oscillation is in the kilohertz range. It can penetrate the vehicle sound systems – either directly (conductive) or inductive – where the ripple is heard as a howl in the loudspeaker system.

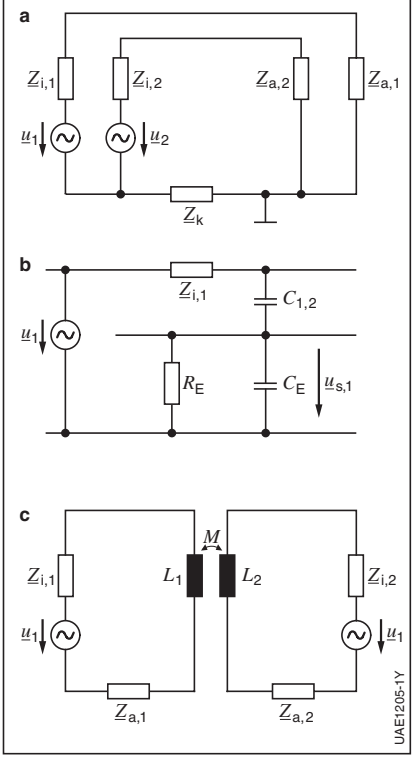
Pulses in the vehicle electrical system

Voltage pulses are generated on the supply lines when electrical equipment is switched on and off. They are routed to adjacent systems directly through the power supply (through conductive coupling, Figure 3a) and indirectly by coupling from connecting lines (through inductive and capacitive coupling, Figure 3b and Figure 3c). These unwanted interference pulses can cause anything from malfunction to total destruction of the adjacent systems. The signal shapes and amplitudes acting as disturbance values depend on the configuration of the vehicle electrical system, such as the ground concept, the position of the wiring harnesses and the individual lines in the wiring harness.

The wide variety of pulses that occur in the vehicle are classified into typical pulse shapes. Major parameters are pulse amplitude, signal rise and fall times as well as internal resistance of the pulse source. A suitable selection of permitted values for interference emission from the sources of interference and the required values for interference immunity ensures that there is no unintended behavior of the electronic systems without unnecessarily high effort for interference suppression. This can be done using various concepts. For example, a higher interference emission for the electric actuators and motors can be allowed than for typical pulse sources if the electronic components, such as control units and sensors, are designed with corresponding interference resistance.

Figure 3: Conductive, capacitive and inductive coupling

a) Conductive coupling of interference signals,
b) Capacitive coupling of interference signals,
c) Inductive coupling of interference signals.
$\underline{Z}_i$ Internal resistance (complex size),
$\underline{Z}_a$ Terminating resistance (complex size),
$\underline{u}_1$, $\underline{u}_2$ Voltage sources (complex size),
$\underline{u}_s$ Interference voltage (complex size),
R_E Input resistance,
C_E Input capacitance,
$C_{1,2}$ Capacitance between both conductors,
L_1, L_2 Inductance of the conductors,
M Inductive coupling,
Z_K Coupling impedance.



Potentially susceptible devices
Electronic control units and sensors are potentially susceptible devices for interference signals that enter the vehicle electrical system from the outside (Figure 2). These interference signals come from the neighboring systems in the vehicle. Malfunctions occur when the system loses the ability to distinguish between interference signals and useful signals. The possibility of taking effective measures depends on the characteristics of useful and interference signals.

The control unit is unable to distinguish between useful and interference signals if the characteristic of an interference signal is similar to that of a useful signal – for instance, when a pulse-shaped interference signal is at the same frequency as the signal from a wheel-speed sensor. Critical frequencies are those in the range of useful signal frequencies ($f_S \approx f_N$) and a low multiple of a signal frequency.

Special feature in vehicle electrical systems with high voltages
High voltage and currents in motor vehicles with hybrid and electric drives pose new challenges. In addition to the contact protection required due to the high voltages, EMC is particularly important. In the components of the power electronics (e.g. inverter and DC/DC converter) and in the electrical machine, large interference signals can arise. Only by using shielding concepts along with suitable filter circuits can these interference signals be reduced enough to allow compliance with the required limit values for mobile communication in the vehicle.

Another special feature is the charging devices with their connections to the public power supply network. This interface, and the resulting requirements with regard to permitted interference emission into the public power grid and the necessary interference protection, must be taken into consideration in the overall design of vehicle electrical systems with high voltages.

High-frequency vibrations and oscillations in the vehicle electrical system
In addition to the low-frequency ripple in vehicle electrical systems and the pulses in these systems, high-frequency vibrations and oscillations are caused in many electromechanical and electronic components by switching operations. Examples include commutation operations in brush motors or in electronically commutated electric motors, activation of output stages or circuits in digital electronics (such as the CPU core of an electronic control unit). These oscillations can spread – more or less attenuated – into the vehicle electrical system via the connected lines, particularly the power supply lines, and can get into the systems for communications electronics via the power supply lines or the capacitive and inductive coupling in the wiring harness.

Depending on whether the measured spectrum of radio interference voltage has a more or less continuous curve or consists of discrete lines, we refer to it as broad-band interference or narrow-band interference. Broad-band interference is caused, for example, by electric motors in wiper drives, control systems, fans, fuel pumps or alternators, but also by certain electronic components. Narrow-band interference is caused, for example, by electronic control units with microprocessors. This allocation depends on the used bandwidth of the radio service under consideration or the bandwidth used during the measurement compared to the characteristics of the interference signal. The various interference signals also have various effects on analog and digital radio systems.

Disturbances can travel through the electric lines in the wiring harness to the supply and signal connections and into the connecting cables of the antennas. They can also be received by the antennas directly and, just like the desired useful signals, they can get into the receiver circuits of the mobile communication devices. These high-frequency disturbances can interfere with vehicle communication systems, because they are often found in the frequency and amplitude range of the useful signal. Narrow-band interference is

Figure 4: Time signal and spectrum for a trapezoidal pulse

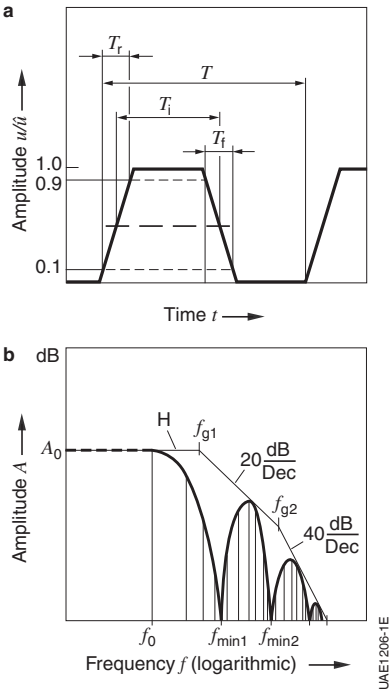
a) Dependency on time,
b) Dependency on frequency.
T Period duration,
T_r Rise time (from 10 % to 90 %),
T_f Release time (from 90 % to 10 %),
T_i Pulse duration,
A_0 Amplitude,
$\hat{u}$ Pulse amplitude,
f_0 Basic frequency of the time signal,
f_{n-1} Harmonics,
f_{min} Periodic minimums,
f_g Fundamental frequencies,
k Pulse duty factor,
n Number of the nth harmonic,
m Number of lines between minimums,
H Envelope,
Dec Decade.

$$f_0 = \frac{1}{T}, \qquad k = \frac{T}{T_i},$$

$$f_{n-1} = \frac{n}{T} = n f_0, \qquad f_{min} = \frac{n}{T_i},$$

$$f_{g1} = \frac{1}{\pi T_i}, \qquad f_{g2} = \frac{1}{\pi T_r},$$

$$m = \frac{T}{T_i}, \qquad A_0 = 2\hat{u}\,T_i.$$



particularly critical, because it has signal characteristics that are very similar to the spectrum of transmitters (Figure 4).

Electrostatic discharge

The subject of potential danger to components and electronic circuits from electrostatic discharge (ESD) also belongs to the field of EMC. The task here is to protect components and equipment from interference or damage by static discharge from humans or from machinery (exposed components) during operation, production and maintenance. This involves both adopting appropriate methods for handling equipment and designing equipment in such a way that the voltages (up to several thousand volts) produced by electrostatic discharges are reduced to acceptable levels.

Immunity to interference from electromagnetic fields

Sources of interference and potentially susceptible devices

During operation, vehicles and their electronic equipment are exposed to various electromagnetic waves, for example from radio, television and wireless transmitters, which are operated at a fixed location, or even from within the same or a neighboring vehicle. Electronic circuits must not experience interference from the electromagnetic fields of the transmitters and the resulting, unwanted voltages and currents.

The large number of electric lines in the wiring harness as well as internal structures in devices (such as printed circuit boards in electronic control units or design technology and joining and bonding techniques in actuators) function as an antenna structure. Depending on their geometric dimensions and the electromagnetic wave frequencies, they receive transmitter signals and conduct them to the semiconductor components. Unmodulated and modulated high-frequency signals can be demodulated at the pn transitions in the semiconductor components. This can lead to level shifts caused by the direct component or to the superposition of transient interference signals as a result of demodulated LF components of the interference signal. The car-

rier frequency $f_{S,HF}$ is usually much higher than the useful-signal frequencies f_N. LF components of the interference signal are particularly critical if its frequency $f_{S,NF}$ is in the range of the useful-signal frequencies f_N. Interference signals with frequencies far lower than those of the useful signals can lead to disturbances due to intermodulations.

Electronic components must be designed in such a way that their function is not impaired by the externally generated interference signals in the electronic circuits.

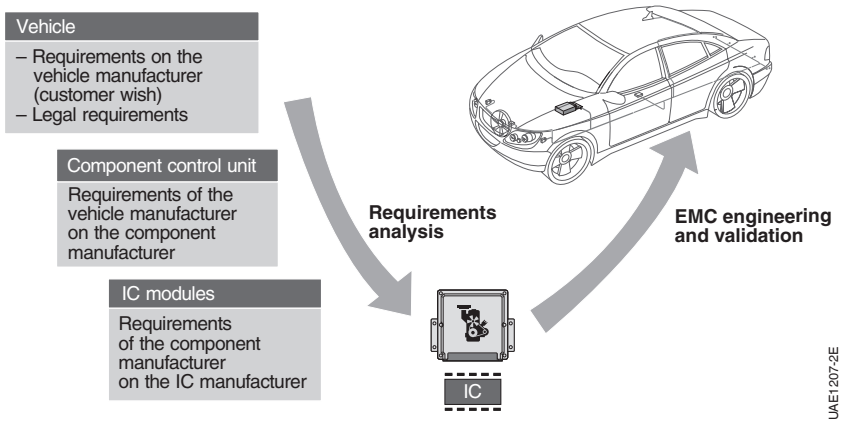
EMC-oriented development

The importance of EMC-oriented development in automotive electronics is increasing at all levels (see V model). Nowadays, safeguarding of EMC in vehicles influences the design of semiconductor components and of modules (including suitable design technology and the joining and bonding techniques of integrated circuits for components and systems) as well as the configuration of the entire vehicle. Thus EMC-oriented design is an integral part of a state-of-the-art development process at all levels. Today, retroactively suppressing interference is usually no longer possible or, at the very least, extremely time consuming and costly.

EMC requirements analysis

When development of a new electronic system for a motor vehicle begins, necessary requirements must first be analyzed and codified in the customer specifications (Figure 5). In addition, the vehicle manufacturer compiles EMC requirements documents, which take into account legal requirements and expectations of vehicle customers. These EMC specifications contain the requirements for vehicles and the resulting requirements for electric systems and components that are being installed in the vehicles. From these requirements, the system or

Figure 5: Requirements analysis and EMC development based on the V model



component manufacturer derives its own requirements for components and specifies requirements for design elements, electronic wiring and semiconductor components. Finally, IC and component manufacturers derive from this the requirements for their own products.

EMC development and verification

During the development stages, these requirements must be defined immediately at all levels in the customer specifications (Figure 5, see also V-model) and taken in to account early on during the vehicle design stage (e.g. the wire harness configuration and its position in the vehicle, power supply and ground concept, and the installing point of electronic components). The requirements derived from the vehicle must be taken into account, keeping in mind the system and component design (e.g. system topology, circuit design, housing design, circuit board layout, design and joining and bonding techniques). Similarly, the component manufacturer takes into account the EMC requirements during the IC and filter design stages. During the development process, the effectiveness of individual development steps is verified, in which various prototypes are metrologically assessed or variant studies using numerical simulation methods are implemented.

EMC validation

During the closing stages of EMC development, the requirements set out in the functional specifications and resulting customer specifications are validated in order to prove their worth. This validation is typically carried out using standardized test methods in accordance with legal mandates, standards and EMC specifications for component and vehicle manufacturers. The measurements to be carried out, operating statuses used and limit values to be observed are written down in test plans and the results are documented in a qualification report.

EMC measuring techniques

EMC measuring techniques are an important tool with regard to EMC-compliant development. Through the use of suitable test methods, the effectiveness of design standards like a selection of suitable semiconductor components, circuit design, circuit board layout, design and housing design are drafted and checked. On the other hand, the EMC test methods used in assessing are conducive to complying with EMC and legal requirements with regard to the release of components and vehicles.

A wide variety of test methods is used for testing interference immunity and interference emission. Depending on the methods used to assess interference phenomena, they can be roughly divided into methods operating in the time domain (pulse generators, oscilloscopes), and those operating in the frequency domain (sine-wave generators, test receivers, spectrum analyzers).

In measuring technology, interference signals are specified as relative quantities of interference emission in dB (decibels). The values for interference suppression (pulse amplitude, transmitter field strengths) are typically specified directly (Table 1).

Table 1: Measured values

Physical quantity	Reference variable	Unit	Calculations
Interference emission			
Voltage L_U	1 µV	dB(µV)	L_U = 20 lg (U/1 µV)
Current L_I	1 µA	dB(µA)	L_I = 20 lg (I/1 µA)
Field strength L_E	1 µV/m	dB(µV/m)	L_E = 20 lg(E/1 µV/m)
Power L_P	1 mW	dB(mW)	L_P = 10 lg (P/1 mW)
Interference immunity			
Voltage U	–	V	–
Current I	–	A	–
Field strength E	–	V/m	–
Field strength H	–	A/m	–
Power P	–	W	–

EMC test methods

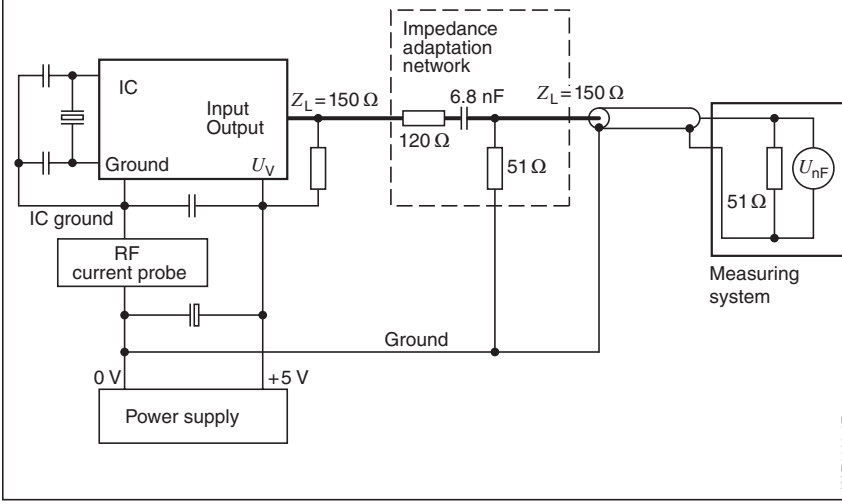
EMC test methods are described in relevant standards and divided into methods for the entire vehicle, for the components and systems (e.g. control units, sensors and actuators) and for integrated circuits (IC) and modules.

IC test methods
During the metrological assessment of integrated circuits (IC), the methods used are designed such that only the component itself is assessed instead of the combination of components with a peripheral circuit and larger conductor structures. The goal is to find out more about the EMC peformance of an IC, independent of its various applications, e.g. in order to compare different ICs of the same type. The standardized test methods for this are divided into conducted and irradiated test methods for interference emission (IEC 61967, see Table 3), interference immunity with respect to electromagnetic fields (IEC 62132, see Table 3) and test methods for sending and influencing pulses (IEC 62215, see Table 3) and ESD (electrostatic discharges). Figure 6 shows an example of a test circuit measuring conducted disturbances on individual IC pins.

Component measurement methods
In assessing the devices in the laboratory, conducted test methods involving radiation are used. Test objects are always operated under standardized conditions. The power is supplied via vehicle line impedance stabilization networks (LISN), which simulate a uniform wiring harness. The measurement configuration is typically completed on a laboratory table with a ground plane. The test object is connected to measurement peripherals, which simulate its functioning in a realistic manner. In order to ensure isolation from its surroundings, the test setup is run in a shielded room.

Conducted disturbances
High-frequency interference voltages at the supply lines are capacitively decoupled in the Line Impedance Stabilisation Network, and interference currents are measured using suitable current measurement coils (CISPR 25, see Table 2).

Figure 6: IC measurements, line-bound interference emission (IEC 61967-4)



Immunity to pulsed interference is tested using special pulse generators which produce standardized test impulses in accordance with ISO 7637-2, Table 2. The coupling of pulsed interference on signal and control lines is reproduced using capacitive coupling clamps in accordance with ISO 7637-3. As with high-frequency interference emission measurements, the interference emission of pulses is measured in a standardized test setup through the use of relevant switches and oscilloscopes.

High-frequency interference coupling and interference emission

Interference immunity test

Coupling of electromagnetic waves for component measurements is achieved through the use of TEM waveguides like the stripline and the TEM cell (TEM, transverse electromagnetic mode) or the power coupling procedure BCI (Bulk Current Injection) and irradiation with antennas.

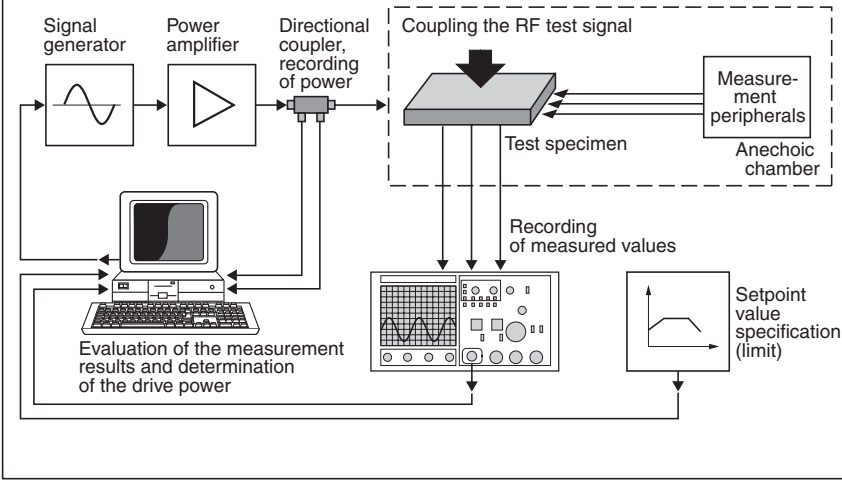
The principal measurement configuration always consists of a coupler for high frequencies in a shielded measuring room, a test object with a wiring harness, measurement peripherals, and devices that produce high frequencies and recording and processing measured values (Figure 7).

In the case of a stripline, (ISO 11452-5, Table 2) the wiring harness is arranged in line with the direction of propagation of the electromagnetic wave between a strip conductor and a base plate. When a TEM cell is used (ISO 11452-3), the test object and a section of the wiring harness are arranged at a right angle to the propagation direction of the electromagnetic waves. The BCI method (ISO 11452-4) involves superimposing an RF current (radio frequency) on the wiring harness by means of a current clamp.

When the test setup is irradiated, the transmitting antennas are set up at various locations. Through this process, both the coupling of electromagnetic fields in the wiring harness and the coupling in the test object are reproduced.

These and further test methods, such as one for testing low-frequency magnetic fields (ISO11452-8) and test methods for reproducing the interference coupling of mobile transmitters within close range (ISO 11452-9), are described in various portions of the ISO 11452 (Table 2).

Figure 7: Basic design of interference immunity tests



Measuring interference emission
The measuring principles for the immunity test can essentially be used to measure interference emission as well (CISPR 25). The TEM waveguides, current clamps and antennas function as receiving elements for the interference emitted by the test objects. For measuring the interference, the test receivers are connected directly to the receiving measuring instruments (Figure 8).

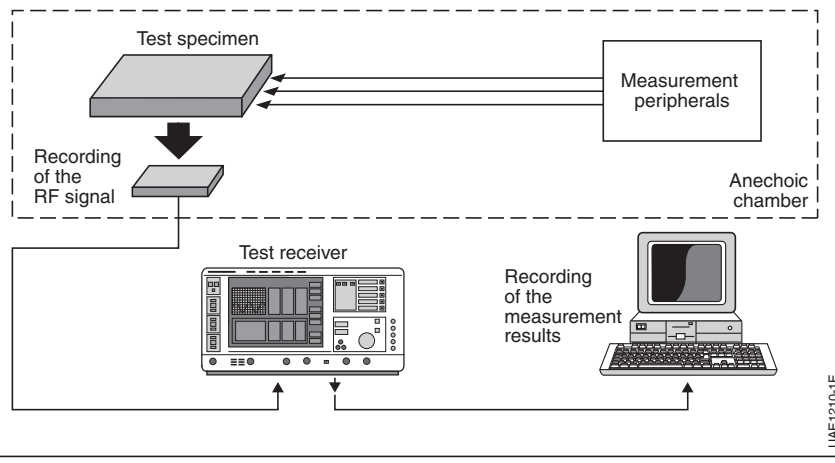
To decouple from the environment – that is, to ensure that only the interference produced by the test specimen is actually measured during interference emission measurements – and to minimize the emission of high-frequency signals in the environment when interference immunity tests are in progress, the testing for high-frequency disturbance values is carried out in electromagnetically shielded rooms. In order to avoid reflections and room resonance, shielded rooms are lined with high frequency absorbers when radiated signals are used.

Test methods for electrostatic discharges
When assessing the interference immunity of ESD, special high voltage impulse generators are used. In these generators, the ESD impulse is reproduced by charging a smoothing capacitor and a targeted discharge via a discharge resistor. The capacity of the smoothing capacitor and the discharge resistor determine the energy output and the pulse shape. With a suitable discharge probe, the ESD pulse is applied to the coupling point, for example a control unit pin, either by a targeted rollover or after contact via a discharge switch in the alternator.

Vehicle testing procedures
Measuring interference immunity
The immunity of electronic systems to electromagnetic fields radiated by high-power transmitters is tested on vehicle level in a special anechoic chambers (Figure 9). Here, high electrical and magnetic field strengths can be generated, and the whole vehicle can be exposed to them (ISO 11451, Table 2).

An interference emission of the entire vehicle is carried out with external antennas, either in standardized open area test sites or in anechoic chambers

Figure 8: Basic design of interference emission measurement

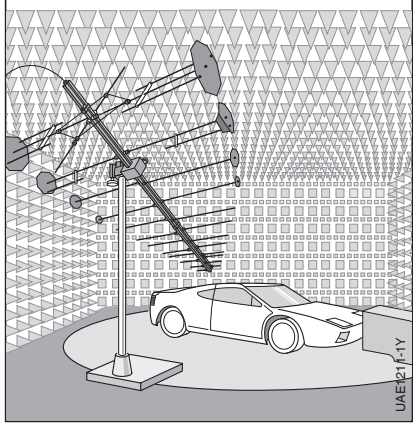


(IEC/CISPR 12, Table 2). The interference effects of the vehicle electrics and electronics on radio reception in the vehicle are measured using highly sensitive test receivers or spectrum analyzers. As far as possible, the original vehicle antenna is left in place, and measurements are taken at the receiver input terminal inside the vehicle in order to collect accurate data. With the help of a suitable test circuit, the impedance of the test receiver is adjusted to the input impedance of the receiver unit. The necessary limit values and measurement parameters like bandwidth and high frequency detectors are derived from the operating parameters of the radio services (IEC/CISPR 25).

Selection of EMC tests

The EMC tests carried out on an electric or electronic device depend on the component's area of application and its internal design. Simple electromechanical devices, which do not contain electronic components, do not have to be tested for interference immunity to electromagnetic fields. On the other hand, a variety of individual tests must be specified and implemented in EMC test plans for components containing electronic parts.

Figure 9: Interference immunity measurement in an EMC anechoic chamber for vehicles



EMC simulation

Nowadays, the application of numerical calculation methods is a method for the EMC development process equivalent to EMC measuring technology. Different calculation methods are employed depending on the problem at hand. What is critical in each case is accurate modeling of the sources of interference, potentially susceptible devices and the coupling path.

If the conducted coupling is dominant or the design of the electrical circuit (for example, filter structures) is to be examined, circuit simulation programs are usually used. If the field-coupled influence is dominant or geometric configurations (such as a metallic structure for shielding the electromagnetic fields or antenna geometry) are to be examined, various methods are used to calculate electromagnetic fields.

Thus the user has to select the method to be used depending on the problem to be solved. The EMC simulation can be used both when developing and assessing individual components, for designing components and systems, and for examining the entire vehicle, for example, to determine the optimum location for an antenna.

A significant advantage of the EMC simulation is that a wide variety of variants (for example, using different components or various geometric configurations) can be examined and compared without having to build and measure a design model for each one.

Legal requirements and standards

EMC type approval

The form of electromagnetic compatibility for motor vehicles is prescribed by law such that, in addition to other requirements (for example, for brakes, light, exhaust gas), requirements concerning interference immunity to electromagnetic fields and regarding the maximum permissible interference emission (interference suppression) also have to be fulfilled for the type approval of vehicles. The currently applicable directive UNECE R10, Revision 5 from 2014 [1] is based on a directive from 1972, which merely contained requirements regarding interference suppression for the vehicle. Since then, this directive has been repeatedly adapted and supplemented to account for technical progress.

The current edition takes into account not only the requirements for conventional motor vehicles, but also requirements for electric and hybrid vehicles with built-in charging devices. Since these vehicles come into contact with the public power grid, there must be assurance of appropriate interference immunity to line-bound interference from the power grid and a corresponding limit to the interference from the vehicle into the power grid.

Two different paths can be taken for the type approval of vehicles. Usually, vehicle manufacturers apply for type approval via type test for the entire vehicle. However, it is also possible to obtain a type approval for electrical and electronic submodules. In this case, in addition to the limit values for interference immunity to electromagnetic fields and the interference radiation, requirements regarding the pulsed interference in the vehicle electrical system also have to be fulfilled.

In addition to regulations governing the type approval procedure for vehicles or components, UNECE R10 also specifies test methods and maximum limits. The test methods are based on the applicable international standards (Table 2); the limit values and precise execution are explicitly specified in the directive.

The defined limit values represent minimum requirements. In practice, just barely complying with the limit values specified is frequently insufficient to ensure interference-free reception or mobile communication in the vehicle. Depending on their vehicle concepts, therefore, the motor vehicle manufacturers specify in their own customer specifications both a raised requirement for interference immunity and a lower interference emission level for protecting radio reception. For the development of electric and electronic components, therefore, these requirements have to be taken into account from the start by consultation between the vehicle manufacturers and suppliers of vehicle electrical systems.

Standards

The individual, different test methods are described in the international ISO and IEC/CISPR standards according to Table 2 for vehicles and their electric/electronic components and systems. The IEC standards according to Table 3 treat the measuring methods for integrated circuits and semiconductor components. Usually the standards do not specify fixed limit values, but limit value classes. Depending on the individual vehicle concepts, this should enable the users of the standards (vehicle manufacturers and suppliers) to provide optimum technical and economical coordination between the requirements regarding interference immunity and interference emission.

Table 2: International standards for motor vehicles and components

Interference immunity	
Designation	Title
Road vehicles – Electrical disturbances from conduction and coupling	
ISO 7637-1	Part 1: Definitions and general considerations
ISO 7637-2	Part 2: Electrical transient conduction along supply lines only
ISO 7637-3	Part 3: Vehicles with nominal 12 V or 24 V supply voltage – Electrical transient transmission by capacitive and inductive coupling via lines other than supply lines
ISO/TR 7673-5	Part 5: Enhanced definitions and verification methods for harmonization of pulse generators according to ISO 7637
Road vehicles – Vehicle test methods for electrical disturbances from narrowband radiated electromagnetic energy	
ISO 11451-1	Part 1: General principles and terminology
ISO 11451-2	Part 2: Off-vehicle radiation sources
ISO 11451-3	Part 3: Onboard transmitter simulation
ISO 11451-4	Part 4: Bulk current injection (BCI)
Road vehicles – Component test methods for electrical disturbances from narrowband radiated electromagnetic energy	
ISO 11452-1	Part 1: General principles and terminology
ISO 11452-2	Part 2: Absorber-lined shielded enclosure
ISO 11452-3	Part 3: Transverse electromagnetic mode (TEM) cell
ISO 11452-4	Part 4: Bulk current injection (BCI)
ISO 11452-5	Part 5: Stripline
ISO 11452-7	Part 7: Direct radio frequency (RF) power injection
ISO 11452-8	Part 8: Immunity to magnetic fields
ISO 11452-9	Part 9: Portable transmitters
ISO 11452-10	Part 10: Immunity to conducted disturbances in the extended audio frequency range
ISO 11452-11	Part 11: Reverberation chamber
ESD – Electrostatic discharge	
ISO 10605	Road vehicles – Test methods for electrical disturbances from electrostatic discharge

Interference emission	
Designation	Title
Vehicles, boats and internal combustion engines	
IEC/CISPR 12	Radio disturbance characteristics – Limits and methods of measurement for the protection of off-board receivers
IEC/CISPR 25	Limits and methods of measurement for the protection of on-board receivers

Table 3: International standards for semiconductor components (IC)

Pulses

Designation	Title
Integrated circuits – Measurement of impulse immunity	
IEC/TS 62215-2	Part 2: Synchronous transient injection method
IEC 62215-3	Part 3: Non-synchronous transient injection method

Interference immunity

Designation	Title
Integrated circuits – Measurement of electromagnetic immunity	
IEC 62132-1	150 kHz to 1 GHz – Part 1: General conditions and definitions
IEC 62132-2	Part 2: Measurement of radiated immunity – TEM cell and wideband TEM cell method
IEC 62132-3	150 kHz to 1 GHz – Part 3: Bulk current injection (BCI) method
IEC 62132-4	150 kHz to 1 GHz – Part 4: Direct RF power injection method
IEC 62132-5	150 kHz to 1 GHz – Part 5: Workbench Faraday cage method
IEC 62132-6	150 kHz to 1 GHz – Part 6: Local injection horn antenna (LIHA) method
IEC 62132-8	Part 8: Measurement of radiated immunity – IC stripline method
IEC/TS 62132-9	Part 9: Measurement of radiated immunity – Surface scan method

Interference emission

Designation	Title
Integrated circuits – Measurement of electromagnetic emissions	
IEC 61967-1	150 kHz to 1 GHz – Part 1: General conditions and definitions
IEC/TR 61967-1-1	Part 1: General conditions and definitions – Near-field scan data exchange format
IEC 61967-2	150 kHz to 1 GHz – Part 2: Measurement of radiated emissions – TEM cell and wideband TEM cell method
IEC/TS 61967-3	150 kHz to 1 GHz – Part 3: Measurement of radiated emissions – Surface scan method
IEC 61967-4	150 kHz to 1 GHz – Part 4: Measurement of conducted emissions; 1 Ohm/150 Ohm direct coupling method
IEC/TR 61967-4-1	150 kHz to 1 GHz – Part 4-1: Measurement of conducted emissions – 1 Ohm/150 Ohm direct coupling method – Application guidance to IEC 61967-4
IEC 61967-5	150 kHz to 1 GHz – Part 5: Measurement of conducted emissions; Workbench Faraday cage method
IEC 61967-6	150 kHz to 1 GHz – Part 6: Measurement of conducted emissions – Magnetic probe method
IEC 61967-8	Part 8: Measurement of radiated emissions – IC stripline method

Additional references

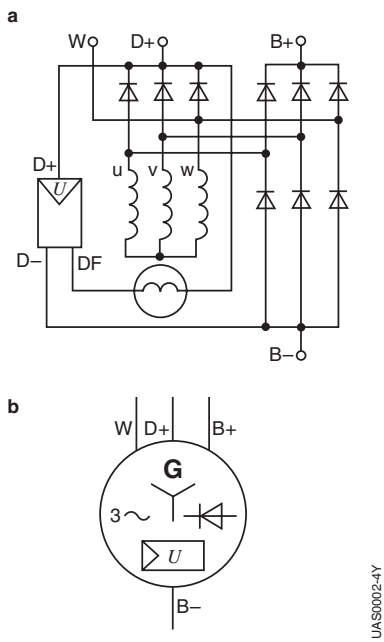
[1] UNECE R10, Revision 5: Uniform provisions concerning the approval of vehicles with regard to electromagnetic compatibility

Symbols and circuit diagrams

The electrical systems in motor vehicles contain a large number of electric and electronic devices for controlling and managing the engine, and for safety and comfort and convenience systems. An overview of the complex vehicle electrical system circuits is only possible with meaningful symbols and circuit diagrams. Circuit diagrams such as schematic diagrams and terminal diagrams help in troubleshooting, simplify the installation of additional devices, and facilitate fault-free connection when retrofitting or altering the electrical equipment in motor vehicles.

Figure 1: Circuit diagram and circuit symbol of an alternator with regulator
The circuit symbols contains, in addition to the symbol for the alternator/generator G, the symbols for the three windings (phases), the star connection, the diodes, and the regulator.

a) With internal circuitry,
b) Circuit symbol.



Circuit symbols

Standards

The circuit symbols shown in Table 1 are a selection of standardized circuit symbols which are suitable for automotive electrics. But for a few exceptions, they correspond to the standards of the International Electrotechnical Commission (IEC).

The European Standard EN 60617 (Graphical symbols for diagrams, [1]) corresponds to the International Standard IEC 60617. It exists in three official versions (German, English and French). The standard contains symbol elements, identifiers and above all circuit symbols for the following areas: General applications (Part 2), Conductors and connecting devices (Part 3), Basic passive components (Part 4), Semiconductors and electron tubes (Part 5), Production and conversion of electrical energy (Part 6), Switchgear, controlgear and protective devices (Part 7), Measuring instruments, lamps and signalling devices (Part 8), Telecommunications: Switching and peripheral equipment (Part 9), Telecommunications: Transmission (Part 10), Architectural and topographical installation plans and diagrams (Part 11), Binary logic elements (Part 12) and Analogue elements (Part 13).

Requirements

Circuit symbols are the smallest elements in a circuit diagram and the simplified graphical representation of an electrical device or a part thereof. The circuit symbols show the operating concept of a device and represent in circuit diagrams the functional correlations of a technical sequence. Circuit symbols do not take into consideration the shape and dimensions of the device and the position of the connections on the device. A detached representation in the schematic diagram is possible by abstraction only.

A circuit symbol should have the following properties: It should be easily remembered, easily comprehensible, uncomplicated in its graphic representation, and clearly within a classification group.

Circuit symbols consist of circuit-symbol elements and qualifying symbols. The following are examples of qualifying symbols: letters, numbers, symbols, mathematical signs and symbols, unit symbols, and characteristic curves.

If a circuit diagram becomes too elaborate due to the representation of the internal circuitry of a device (Figure 1a) or if not all the details of the circuit are needed to identify the function of the device, the circuit diagram for this special device can be replaced by a single circuit symbol (without internal circuitry) (Figure 1b).

In the case of integrated circuits which demonstrate a high degree of economy of space (this is synonymous with large scale integration of functions in a component), a simplified circuit representation is preferred.

Representation

The circuit symbols are shown without the effect of a physical quantity, i.e. in a de-energized and mechanically non-actuated state. An operating state of a circuit symbol that deviates from this standard representation (basic status) is denoted by an adjacent double arrow (Figure 2).

Circuit symbols and connecting lines (representing electric lines and mechanical linkages) have the same line width.

To avoid unnecessary kinks and crossings of the connecting lines, it is possible to rotate circuit symbols in 90° increments or show a mirror image of them, provided this does not alter their meaning. The direction of continuing lines can be freely selected. Exceptions are the circuit symbols for resistors (connection symbols are only permitted here on the narrow sides) and connections for electromechanical drives (connection symbols are only permitted here on the wide sides, Figure 3).

Junctions are represented both with and without a dot. Where there are crossings without a dot, there is no electrical connection. Connecting points on devices are for the most part not specifically represented. Connecting point, plug, jack or screwed connections are identified by a circuit symbols only at the points required for installation and removal. Other connecting points are identified as standard by dots.

Contact elements with a common drive are identified in an assembled representation in such a way that on actuation they follow a direction of motion which is established by the mechanical linkage (– – –) (Figure 4).

Figure 2: Operating state of the circuit symbol deviating from the basic status
a) NO contact,
b) Electromechanical drive.
1 Basic status,
2 Deviation.

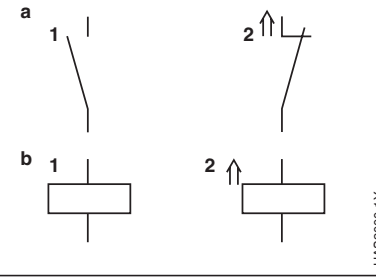


Figure 3: Terminals
a) Resistor,
b) Electromechanical drive.

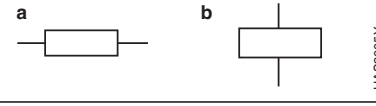


Figure 4: Mechanical linkage at a multi-position switch
0, 1, 2 Switch positions,
15, 30, 50 Terminal designations.

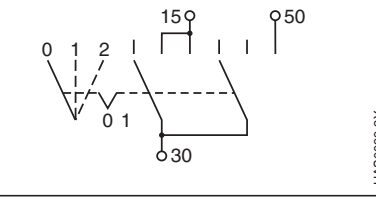


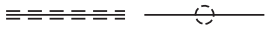
Table 1: Selection of circuit symbols as per EN 60617

Connections

Line; line intersection, without/with a connection

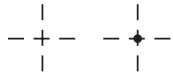


Shielded line



Mechanical linkage; electrical conductor (laid at later stage)

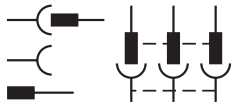
Crossovers (without/with connections)



Connection, general; separable connection (if indication necessary)



Plug connection; socket; plug; three-plug connection

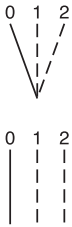


Ground (housing ground, vehicle ground)

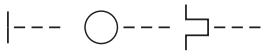


Mechanical function

Switch positions (home position: solid line)



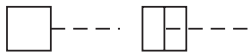
Activation is manual, by sensor (cam), thermal (bimetal)



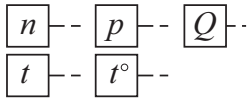
Detent; non-automatic/ automatic return in direction of arrow (button)



Actuator, general (mech., pneum., hydraul.); piston actuation



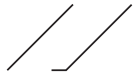
Actuation by rotational speed n, pressure p, quantity Q, time t, temp. $t°$



Variability/adjustability, not intrinsic (external), general



Variability/adjustability, intrinsic, caused by applied physical variable, linear/non-linear



Variability/adjustability, general



Switches

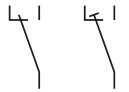
Pushbutton switch, NO/NC contact



Detent switch, NO/NC contact



Changeover contact, non-bridging/bridging



UAS1245-1E

Switches

Two-way normally open contact with three positions (e.g. turn-signal indicator switch)



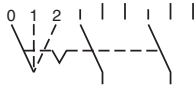
NO/NC contact



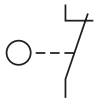
Contact with two makes



Multiple-position switch



Cam-operated switch (e.g. contact breaker)



Thermostatic switch

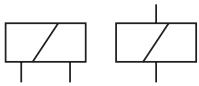


Release/trip device



Various components

Actuators with one winding



Actuator with two windings acting in same direction



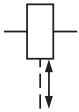
Actuator with two windings acting in opposing directions



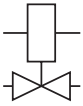
Electrothermal actuator, thermal relay



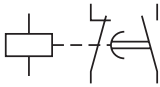
Electrothermal actuator, linear solenoid



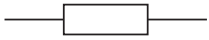
Solenoid valve, closed



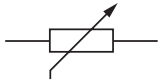
Relay (actuator and switch), example: non-delayed-break NC contact and delayed-break NO contact



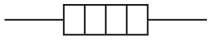
Resistor



Potentiometer (with three connections)



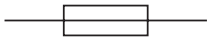
Heating resistor, glow plug, flame plug, heated window



Antenna



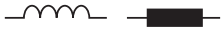
Fuse



Permanent magnets



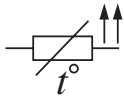
Winding, inductive



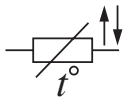
UAS1246E

Various components

PTC resistor



NTC resistor

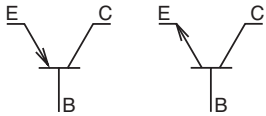


Diode, general, current in direction of triangle tip

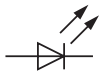


PNP transistor
NPN transistor

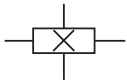
E = Emitter (arrow points in direction of flow)
C = Collector, positive
B = Base (horizontal), negative



Light-emitting diode (LED)



Hall generator

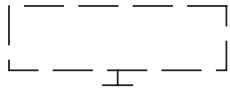


Devices in motor vehicle

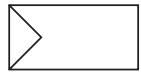
Dotted/dashed line used to delineate or group together associated circuit sections



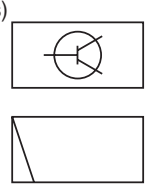
Shielded device, dashed line connected to ground



Regulator, general



Electronic control units (ECUs)



Indicating instrument, general; voltmeter; clock



Rotational-speed indicator; temperature indicator; linear-speed indicator



Battery



Plug-and-socket connection



Light, headlamp



Horn, fanfare horn



Heated rear window (general heating resistor)



Switch, general, without indicator lamp



Switch, general, with indicator lamp



UAS1247E

Devices in motor vehicle

Pressure switch



Relay, general



Solenoid valve, fuel injector, cold-start valve



Thermo-time switch



Throttle-valve switch



Rotary actuator



Auxiliary-air valve with electrothermal actuator



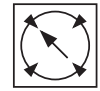
Spark plug



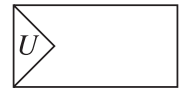
Ignition coil



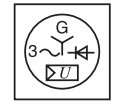
Ignition distributor, general

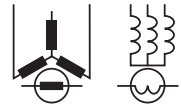


Voltage regulator



Alternator with regulator (without/with internal circuitry)





Electric fuel pump, engine drive for hydraulic pump

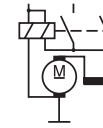


Motor with blower, fan



Starter motor with engagement relay (without/with internal circuitry)



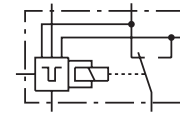


Wiper motor (one/two wiper speeds)





Intermittent-wiper relay



Car radio



UAS1248E

Devices in motor vehicle

Loudspeaker



Piezoelectric sensor



Linear-speed sensor



Voltage stabilizer, stabilizer



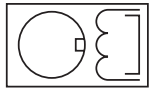
Resistance sensor



ABS rotational-speed sensor



Inductive sensor, coritrolled with reference mark



Air-flow sensor



Hall sensor



Turn-signal flasher, pulse generator, intermittent relay



Air-mass meter



Converter (rate, voltage)



Lambda oxygen sensor (unheated/heated)





Flow-quantity sensor, fuel-level sensor



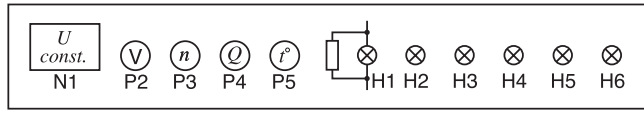
Temperature switch, temperature sensor



Inductive sensor



Instrument-cluster device (dashboard)



UAS1249E

Circuit diagrams

Circuit diagrams are idealized representations of electrical devices, rendered in the form of symbols. Such diagrams also include illustrations and simplified design drawings as needed. A circuit diagram illustrates the relationship between the various devices and shows how they are connected to each other. A circuit diagram may be supplemented by tables, graphs, or descriptions. The type of circuit diagram actually used is determined by its particular purpose (e.g. illustrating the operation of a system), and by the way in which the circuit is represented (Figure 5).

For a circuit diagram to be "legible", it must satisfy the following requirements:

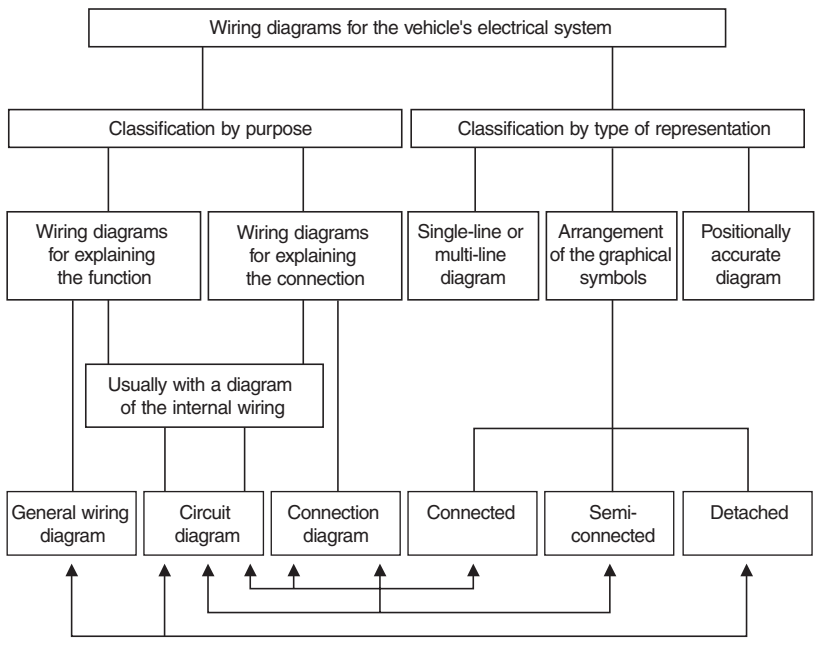
– It must comply with the requirements of the appropriate standards, and deviations must be explained.
– Current paths should be arranged so that current flow or mechanical action takes place from left to right or from top to bottom.

In automotive electrics, block diagrams with single inputs and outputs and with internal circuitry omitted provide a swift overview of the function of a system or a device. The schematic diagram in different methods of representation (layout of the circuit symbols) is the detailed representation of a circuit to identify the function and to carry out repairs. The terminal diagram (with terminal locations of the devices) is used by the Aftersales Service when replacing or retrofitting devices.

According to the method of representation, a distinction is made between

– single- or multi-line representation, and (according to the layout of the circuit symbols)
– assembled, semi-assembled, detached and topographical representation, which can be combined in one and the same circuit diagram.

Figure 5: Classification of the circuit diagrams (in accordance with EN 81346, Part 1 [2])



Block diagram

The block diagram is the simplified representation of a circuit, where only the essential parts are taken into consideration (Figure 6). It is intended to provide a quick overview of the task, design, structure and function of an electrical system or a part thereof, and to serve as a guide for more detailed circuit documentation (schematic diagram).

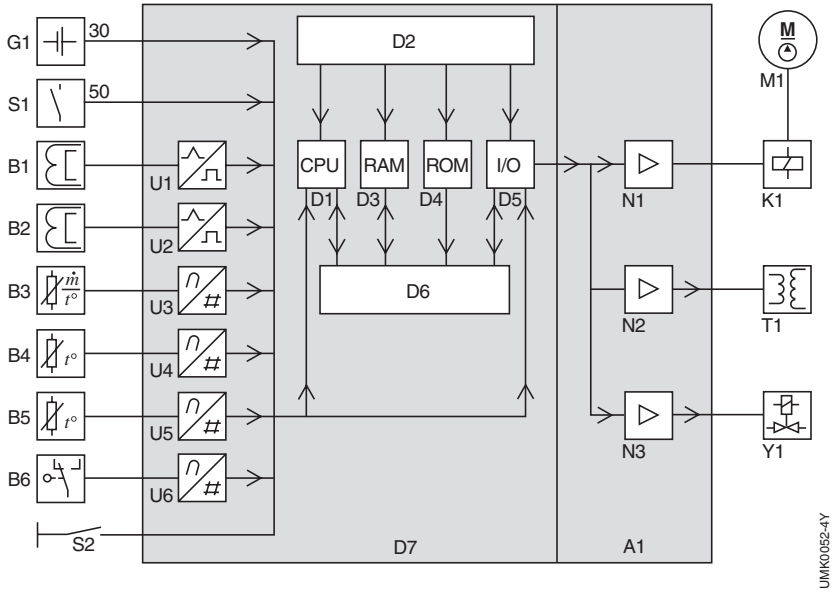
The devices are depicted by squares, rectangles and circles with qualifying symbols inside them similar to EN 60617, Part 2. The lines are for the most part drawn as single lines.

Schematic diagram

A schematic diagram is the detailed representation of a circuit. By clearly depicting individual current paths, it also indicates how an electrical circuit operates. In a schematic diagram, the clear presentation of the circuit's operation, which makes the diagram easy to read, must not be interfered with by the presentation of the individual circuit components and their spatial relationships. Figure 7 shows the schematic diagram of a starter motor in assembled and detached representations.

Figure 6: Block diagram of a Motronic ECU as an example

A1 ECU,
B1 Sensor for engine speed, B2 Sensor for reference mark,
B3 Sensor for air mass, B4 Sensor for intake-air temperature,
B5 Sensor for engine temperature, B6 Throttle-valve switch,
D1 Computing unit (CPU), D2 Address bus, D3 Main memory (RAM),
D4 Program data memory (ROM), D5 Input – Output,
D6 Data bus, D7 Microcomputer,
G1 Battery, K1 Pump relay, M1 Electric fuel pump,
N1 to N3 Driver stages,
S1 Ignition/starter switch, S2 Map changeover switch,
T1 Ignition coil, U1, U2 Pulse shapers, U3 to U6 Analog-digital converters,
Y1 Fuel injector.



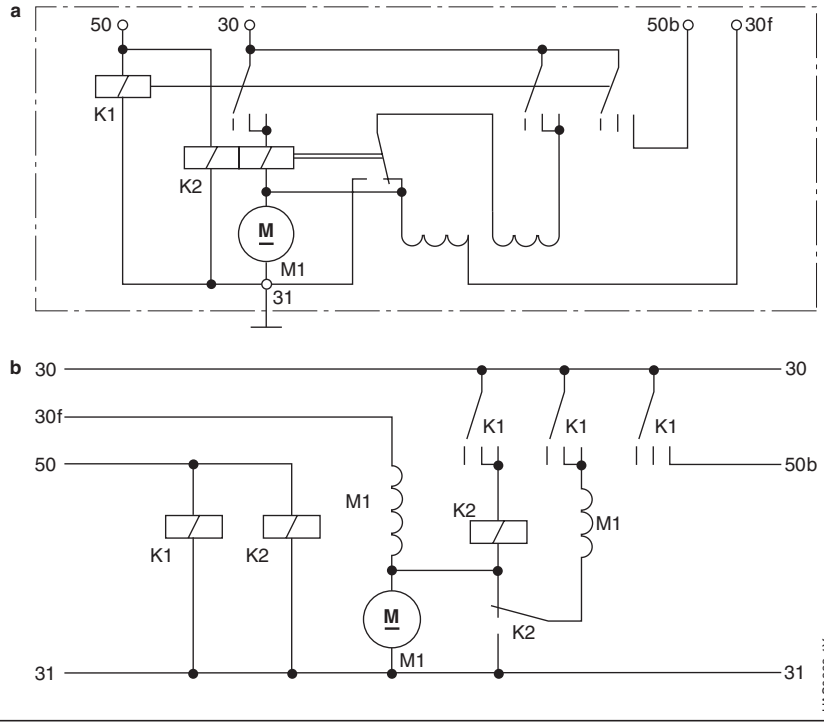
The schematic diagram must contain:
– circuit,
– device identification (EN 81346, Part 2, [2]), and
– pin designation and terminal designation (DIN 72552, [3]).

The schematic diagram can contain:
– A complete representation with internal circuitry in order to facilitate testing, fault localization, maintenance and replacement (retrofitting).
– Reference designations to better locate circuit symbols and destinations, particularly in a detached representation.

Representation of the circuit
Mostly multi-line representation is used in the schematic diagram. For the layout of the circuit symbols, there are, in accordance with EN 81346 Part 1, the following methods of representation which can be combined in the same circuit diagram.

Assembled representation
All the parts of a device are directly represented together assembled and connected to each other by a double slash or discontinuous connecting lines to identify the mechanical linkage. This representation can be used for simple, not very com-

Figure 7: Schematic diagram of a KB-type starter motor for parallel operation using two methods of representation
a) Assembled representation,
b) Detached representation.
K1 Control relay, K2 Engagement relay, holding winding and pull-in winding,
M1 Starter motor with series and shunt windings.
30, 30f, 31, 50b Terminal designations.



prehensive circuits without a loss of clarity (Figure 7a).

Detached representation
Circuit symbols of parts of electrical devices are represented separately and arranged in such a way that each current path can be traced as easily as possible (Figure 7b). The spatial association of individual devices or their parts is not taken into consideration. The priority is for the layout of the current paths to be as linear, clear and crossing-free as possible. The main purpose of this representation is to identify the function of a circuit.

The association of the individual parts must be identified using an identification system in accordance with EN 81346, Part 2. The identification associated with a device is located on each individual, separately represented circuit symbol for that device. Devices shown detached must be indicated once complete and assembled at a point in the circuit diagram if this is necessary to understand the circuit.

Topographical representation
In this representation, the position of the circuit symbol entirely or partially corresponds to the spatial position within the device or part.

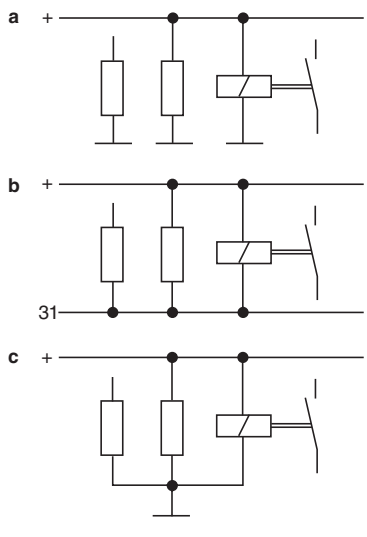
Ground representation
For most motor vehicles, the single-conductor system, in which ground (metal parts of the vehicle) acts as the return conductor, is preferred for its simplicity. If a perfectly conducting connection of the individual ground parts is not guaranteed or if voltages in excess of 42 V are involved, the return conductor is also laid insulated from ground.

All the ground symbols represented in a circuit are electrically connected to each other via device or vehicle ground.

All the devices which contain a ground symbol must be connected to vehicle ground and be electrically conductive.

Figure 8 shows different possibilities of ground representation.

Figure 8: Ground representation
a) Individual ground symbols,
b) Ground continuity,
c) With ground collecting point.
31 Terminal designation.



Current paths and lines
The circuits are arranged in such a way as to produce and clear representation. The individual current paths, with the effective direction preferably from left to right or from top to bottom, should generally run parallel to the edge of the circuit diagram as linearly as possible, crossing-free and without changes of direction.

When there is a bundling of parallel lines, these are grouped into units of three lines each, and there follows a gap to the next group.

Boundary lines and framings
Dot-and-dash separating or framing lines delimit parts of circuits in order to show the functional or design association of the devices or parts.

In automotive electrical systems, this dot-and-dash line represents a non-conductive framing of devices or circuit parts; it does not always correspond to the circuit housing and is not used as device ground. In high-voltage electrical systems, this framing line is often connected to the protective earth conductor (PE),

which is also indicated by a dot-and-dash line.

Break points, identifier, destination designation
Connecting lines (lines and mechanical linkages) which run over a longer distance of the schematic diagram can be interrupted in order to improve clarity. Only the beginning and the end of the connecting line are represented. The association of these break points must be clearly identifiable. The identifier and the destination designation are used for this purpose (Figure 9).

The identifier at associated break points matches. Terminal designations in accordance with DIN 72552 (Figure 9a), the specification of the operating concept and specifications in the form of alphanumeric characters serve as the identifier.

The destination designation is enclosed in parentheses so that it is not confused with the identifier; it consists of the section number of the destination (Figure 9b).

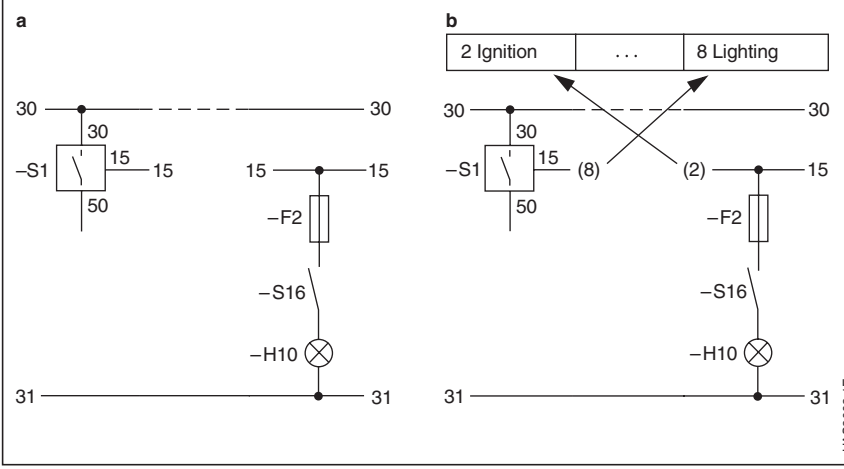
Section designation
The section designation indicated at the top edge of the diagram (formerly called the current path) is used to locate circuit parts. This designation can take three different forms:
– consecutive numbers at the same intervals from left to right (Figure 10a),
– reference to the contents of the circuit sections (Figure 10b),
– or a combination of the two (Figure 10c).

Labeling
Devices, parts or circuit symbols are identified in circuit diagrams with a letter and a number in accordance with EN 81346, Part 2. This identification is placed on the left or underneath the circuit symbol.

The qualifying symbol specified in the standard for the type of devices can be omitted if this does not give rise to ambiguity.

In the case of nested devices, a device is a component part of another, e.g. starter M1 with built-in engagement relay K6. The device designation is then: – M1 – K6.
Designation of associated circuit symbols in a detached representation: Each individual, separately represented circuit symbol of a device is given the designation common to the device.

Figure 9: Identification of break points
a) By terminal designation, e.g. term. 15,
b) By destination designation, e.g. in section 8 and 2.

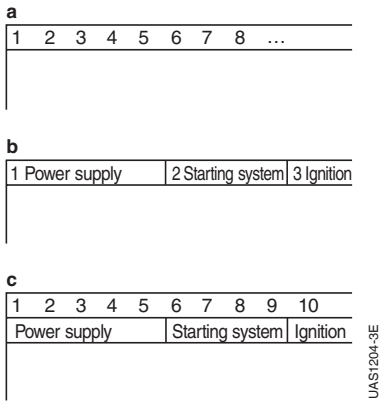


Pin designations (e.g. in accordance with DIN 72552) must be written outside the circuit symbol, and in the case of framing lines preferably outside the framing.

Where the current paths run horizontally: The specifications assigned to the individual circuit symbols are written under the relevant circuit symbols. The terminal designation is located directly outside the actual circuit symbols above the connecting line.

Where the current paths run vertically: The specifications assigned to the individual circuit symbols are written on the left next to the relevant circuit symbols. The terminal designation is located directly outside the actual circuit symbol, on the right if the format is horizontal and on the left next to the connecting line if the format is vertical.

Figure 10: Possible section designations
a) With consecutive numbers,
b) With references to the sections,
c) With a combination of a and b.

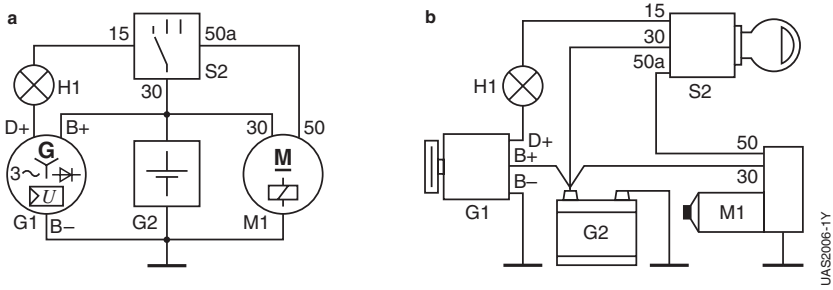


Terminal diagram
The terminal diagram shows the terminal locations of electrical devices and the connected outer and – if necessary – inner conductive connections (lines).

Representation
The individual devices are represented by squares, rectangles, circles and circuit symbols or even pictorially represented, and can be topographically arranged. Circle, dot, plug-in connection or only the led out line are used as terminal points.

The following methods of representation are customary in automotive electrics:
– Assembled, circuit symbols corresponding to EN 60617 (Figure 11a).
– Assembled, pictorial device representation (Figure 11b).
– Detached, device representation with circuit symbols, terminals with destination designations (Figure 12a); lines can be color coded.
– Detached, pictorial device representation, terminals with destination designations (Figure 12b); lines can be color coded.

Figure 11: Example of a terminal diagram (assembled representation)
a) With circuit symbols,
b) With devices.



Labeling
The devices are labeled in accordance with EN 81346, Part 2. Terminals and plug-in connections are identified with the terminal designations present on the device (Figure 11).

In the detached method of representation, the continuous connecting lines between the individual devices are omitted. All outgoing lines from a device are given a destination designation (EN 81346, Part 2), consisting of the identification of the destination device and its terminal designation, and – if necessary – the line color is also given (Figure 13).

Figure 13: Device designation (example: alternator)
a) Device designation (code letter and code number),
b) Terminal designation on device,
c) Device to ground,
d) Destination reference (code letter and code number/terminal designation/wire color).

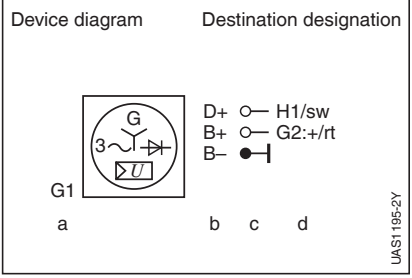
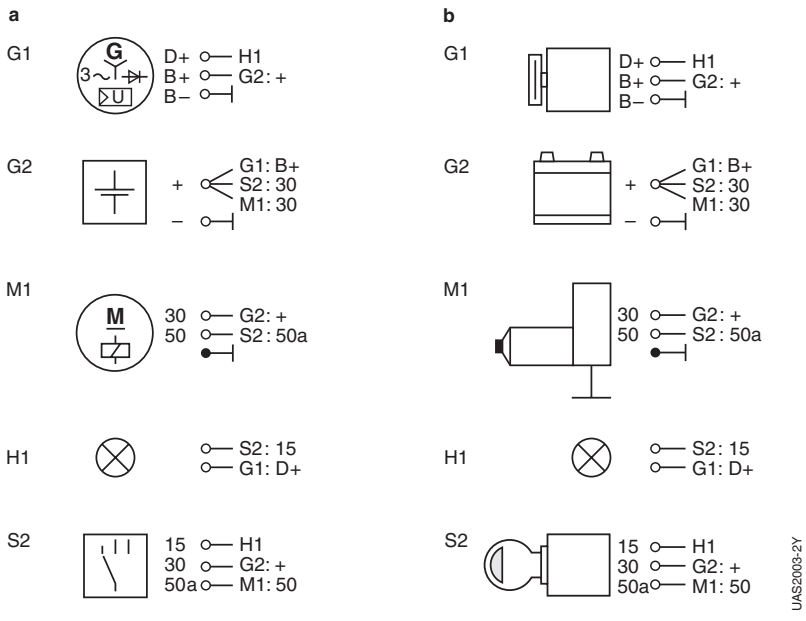


Figure 12: Terminal diagram (detached representation)
a) With circuit symbols and destination designations,
b) With devices and destination designations.
G1 Alternator with regulator, G2 Battery, H1 Alternator indicator lamp,
M1 Starter motor, S2 Ignition/starter switch,
15, 30, 50, 50a Line potential, e.g. terminal 15.



Assembled-representation diagram
For troubleshooting on complex and extensively networked systems with self-diagnosis function, Bosch has developed system-specific schematic diagrams. Bosch makes assembled-representation diagrams for further systems in a great number of motor vehicles available in ESI[tronic] (Bosch Electronic Service Information). This provides automotive repair shops with a useful tool for locating faults or wiring retrofit equipment. Figure 15 shows as an example the assembled-representation diagram for a door locking system.

In contrast to other schematic diagrams, the assembled-representation diagrams use US symbols that are supplemented by additional descriptions (Figure 14). These include component codes – e.g. "A28" (theft-deterrence system) –, which are explained in Table 2 and an explanation of the wire colors (Table 3). Both tables can be called up in ESI[tronic].

Figure 14: Additional descriptions on assembled-representation diagrams
1 Wire color code,
2 Connector number,
3 Pin number (a dashed line between pins indicates that all pins belong to the same connector).

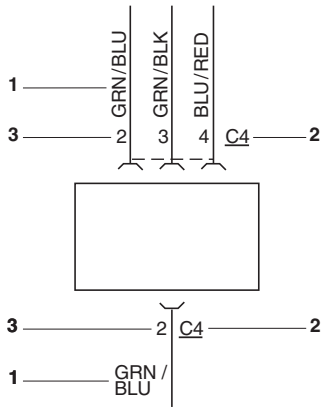


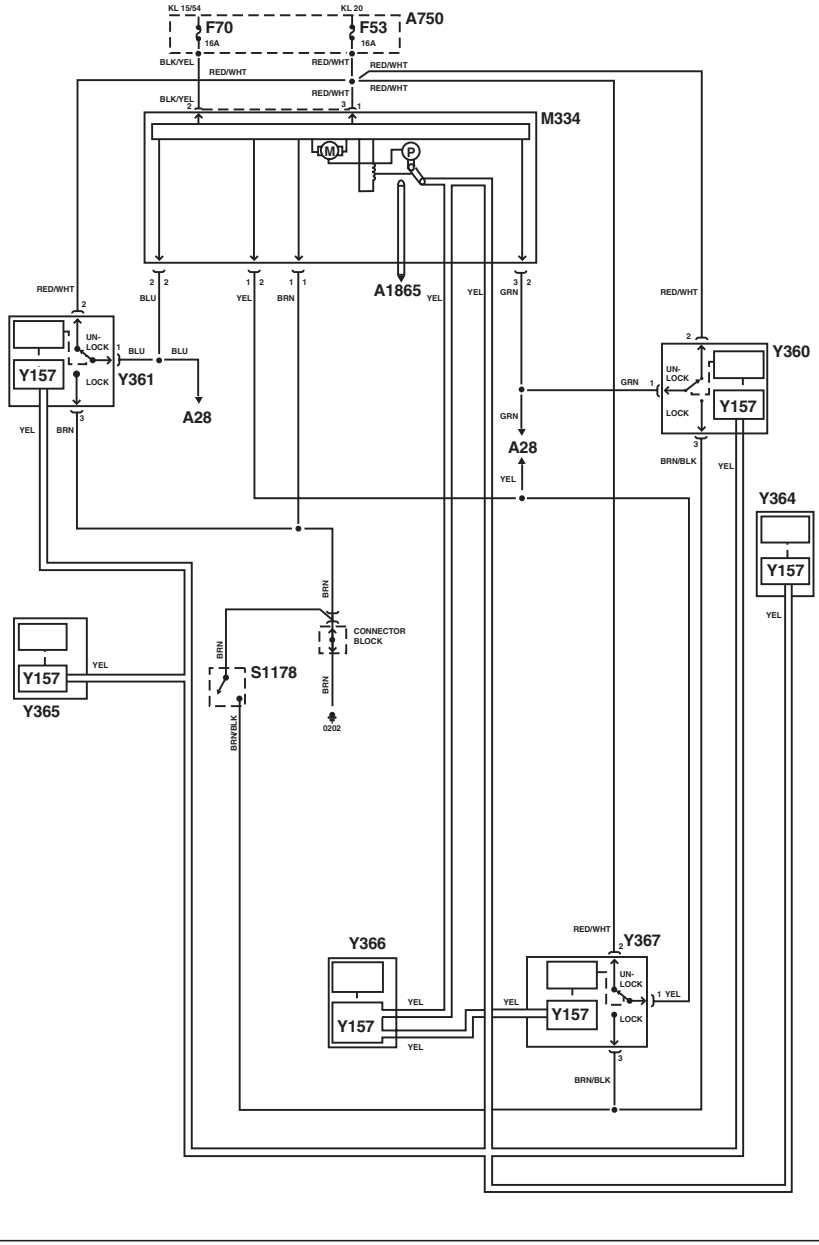
Table 2: Explanation of component codes

Position	Designation
A1865	Electrically adjustable seat system
A28	Theft-deterrence system
A750	Fuse box, relay box
F53	Fuse C
F70	Fuse A
M334	Supply pump
S1178	Warning-buzzer switch
Y157	Vacuum actuator
Y360	Actuator, door, front, right
Y361	Actuator, door, front, left
Y364	Actuator, door, rear, right
Y365	Actuator, door, rear, left
Y366	Actuator, filler cap
Y367	Actuator, lock, luggage compartment, trunk lid, tailgate

Table 3: Explanation of wire colors

Position	Designation
BLK	Black
BLU	Blue
BRN	Brown
CLR	Clear
DK BLU	Dark blue
DK GRN	Dark green
GRN	Green
GRY	Gray
LT BLU	Light blue
LT GRN	Light green
NCA	Color not known
ORG	Orange
PNK	Pink
PPL	Purple
RED	Red
TAN	Tan
VIO	Violet
WHT	White
YEL	Yellow

Figure 15: Assembled-representation diagram of a door locking system (example)

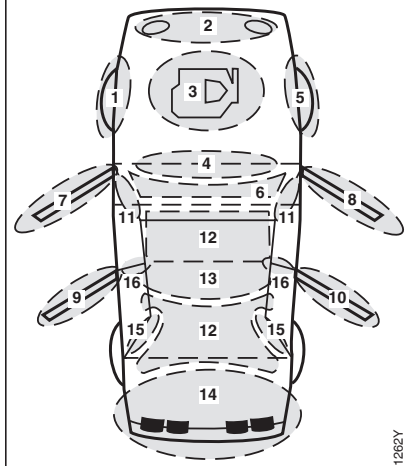


The assembled-representation diagrams are subdivided into system circuits and, if applicable, subsystems (see Table 4). Classification of system circuits reflects the standard ESI[tronic] practice as used for other systems, according to which they are assigned to one of four assembly groups:
– engine,
– body,
– chassis/suspension,
– drivetrain.

Table 4: System circuits

1	Engine management
2	Starting/charging
3	Air conditioner/heating (HVAC)
4	Radiator blower
5	ABS
6	Cruise control
7	Power-window units
8	Central locking system
9	Dashboard
10	Washer/wiper system
11	Headlamps
12	Exterior lights
13	Power supply
14	Grounding
15	Data line
16	Shift lock
17	Theft deterrence
18	Passive safety systems
19	Power antenna
20	Warning system
21	Heated windshield/mirrors
22	Supplementary safety systems
23	Interior lights
24	Power steering
25	Mirror adjuster
26	Soft-top controls
27	Horn
28	Trunk, tailgate
29	Seat adjustment
30	Electronic damping
31	Cigarette lighter, socket
32	Navigation
33	Transmission
34	Active body components
35	Vibration damping
36	Cellular phone
37	Automotive sound system, Hi-Fi
38	Immobilizer

Figure 16: Grounding points
1 Left front fender,
2 Front end,
3 Engine,
4 Bulkhead,
5 Right front fender,
6 Footwell bulkhead/dashboard,
7 Left front door,
8 Right front door,
9 Left rear door,
10 Right rear door,
11 A pillars,
12 Passenger compartment,
13 Roof,
14 Rear end,
15 C pillars,
16 B pillars.



It is important to be aware of the grounding points, particularly when fitting additional accessories. For this reason, ESI[tronic] includes the vehicle-specific location diagram for the grounding points (Figure 16) in addition to the assembled-representation diagrams.

Designations for electrical devices

The designations in accordance with EN 81346 Part 2 (Table 5) are used to provide a clear, internationally comprehensible identification of systems, parts, etc. which are represented by circuit symbols in a circuit diagram. They appear next to the circuit symbols and consist of a series of defined qualifying symbols, letters and numbers (Figure 17).

Figure 17: Device designation

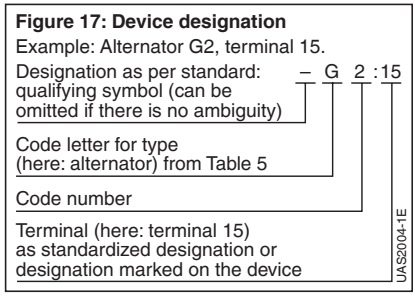


Table 5: Code letters for designating electrical devices as per EN 81346-2

Code letter	Type	Examples
A	System, assembly, component group	ABS ECU, automotive sound system, car radiotelephone, car phone, theft-deterrence system, device assembly, control device, ECU, cruise control
B	Converter of non-electrical into electrical quantities or vice versa	Reference-mark transmitter, pressure switch, fanfare horn, horn, λ oxygen sensor, loudspeaker, air-flow sensor, microphone, oil-pressure switch, sensors of all types, ignition trigger
C	Capacitor	Capacitors of all types
D	Binary element, memory	On-board computer, digital equipment, integrated circuit, pulse counter, magnetic tape recorder
E	Assorted devices and equipment	Heater, air-conditioning system, lamp, headlamp, spark plug, ignition distributor
F	Protective device	Trigger (bimetal), polarity protection device, fuse, current protection circuit
G	Power supply, alternator	Battery, alternator, charger
H	Check device, signaling unit, signaling device	Acoustic signaling unit, function lamp, turn-signal indicator, turn-signal lamp, brake-pad indicator, stop lamp, high-beam indicator, alternator charge indicator, indicator lamp, signaling unit, oil-pressure indicator, optical signaling unit, signal lamp, warning buzzer
K	Relay, contactor	Battery relay, turn-signal flasher, turn-signal relay, engagement relay, starting relay, hazard warning light flasher
L	Inductance	Inductance coil, coil, winding
M	Engine type	Blower motor, fan motor, pump motor for ABS, TCS and ESC hydraulic modulators, windshield-washer and windshield-wiper motors, starter motor, servomotor
N	Regulator, amplifier	Regulator (electronic or electromechanical), voltage stabilizer

Table 5: Code letters for designating electrical devices (continued)

Code letter	Type	Examples
P	Tester	Ammeter, diagnostic socket, rev counter, pressure indication, tachograph, measuring point, check point, speedometer
R	Resistance	Sheathed-element glow plug, flame glow plug, heating resistor, NTC resistor, PTC resistor, potentiometer, regulating resistor, series resistor
S	Switch	Switches and buttons of all types, ignition contact breaker
T	Transformer	Ignition coil, ignition transformer
U	Modulator, converter	DC converter
V	Semiconductor, tube	Darlington, diode, electron tube, rectifier, semiconductors of all types, variable capacitance diode, transistor, thyristor, Zener diode
W	Transmission path, line, antenna	Vehicle antenna, shield part, shielded line, lines of all types, line bundle, ground (continuity) line
X	Terminal, plug, plug-in connection	Terminal stud, electrical terminals of all kinds, spark-plug connector, terminal, terminal strip, electrical cable coupler, line connector, plug, socket, plug connector, (multiple) plug-in connection, distributor connector
Y	Electrically actuated mechanical equipment	Permanent magnet, (solenoid) fuel injector, electromagnetic clutch, electromagnetic brake, electric air slider, electric fuel pump, electromagnet, electric starting valve, transmission control, linear solenoid, kickdown solenoid valve, headlight leveling control, level control valve, switching valve, starting valve, door locking, central locking system, auxiliary-air slider
Z	Electric filter	Screening unit, suppression filter, filter chain, timer

Terminal designations

The purpose of the terminal-designation system for automotive electrical systems specified by the standard (DIN 72552) is to enable the most accurate connection of wires to all the various devices, above all when making repairs and installing spare parts. The terminal designations (Table 6) do not identify the wires because devices with different terminal designations can be connected at the two ends of each wire. For this reason, they need not be written on the wires.

In addition to the terminal designations listed, designations according to DIN VDE standards may also be used on electrical machines. Multiple connectors, for which the number of terminal designations as per DIN 72552 no longer suffice, are numbered by consecutive numbers or letters whose function assignment is not specified by standards.

Table 6: Terminal designations according to DIN 72552

Terminal	Definition
	Ignition coil
1	Low voltage
4	High voltage
4a	From ignition coil I, terminal 4
4b	From ignition coil II, terminal 4
15	Switched positive after battery (ignition-switch output)
15a	Output at the series resistor to the ignition coil and to the starter
	Preheating system
17	Start
19	Preglow
	Battery
30	Line from battery positive terminal (direct)
30a	Battery changeover 12/24 V Line from battery II positive terminal
31	Return wire from battery Negative or ground (direct)
31 b	Return wire to battery Negative or ground via switch or relay (switched negative)
	Battery changeover relay 12/24 V
31a	Return line to battery II negative
31c	Return line to battery I negative
	Electric motors
32	Return line[1])
33	Main terminal[1])
33a	Self-parking
33b	Shunt field
33f	for second reduced-rpm operation
33g	for third reduced-rpm operation
33h	for fourth reduced-rpm operation
33L	Rotation to left (counterclockwise)
33R	Rotation to right (clockwise)

Terminal	Definition
	Starter
45	Separate starter-motor relay, output; starter, input (primary current)
	Dual starters, parallel activation Starting relay for pinion-engagement current
45a	Starter I output, Starter I and II input
45b	Starter II output
48	Terminal on starter and start repeating relay (monitoring the starting process)
	Hazard-warning and turn-signal flashers
49	Input, turn-signal flasher
49a	Output, turn-signal flasher
49b	Output, second turn-signal flasher
49c	Output, third turn-signal flasher
	Starter
50	Starter control (direct)
	Battery changeover relay
50a	Output for starter control
	Starter control
50b	In parallel operation of two starter motors with sequence control
	Starting relay for sequence control of engagement current in parallel operation of two starter motors
50c	Input at starting relay for starter I
50d	Input at starting relay for starter II
	Start-locking relay
50e	Input
50f	Output
	Start repeating relay
50g	Input
50h	Output

[1]) Polarity reversal terminal 32/33 possible

Table 6: Terminal designations according to DIN 72552 (continued)

Terminal	Definition
	Wiper motors
53	Wiper motor, input (+)
53a	Wiper (+), self-parking
53b	Wiper (shunt winding)
53c	Electric windshield-washer pump
53e	Wiper (brake winding)
53i	Wiper motor with permanent magnet and third brush for higher speed)
	Lighting
55	Fog lamps
56	Headlamps
56a	High beam with indicator lamp
56b	Low beam (dipped beam)
56d	Headlamp-flasher contact
57a	Parking lamp
57L	Parking lamp, left
57R	Parking lamp, right
58	Side-marker, tail, license-plate and instrument lamps
58L	left
58R	right
	Alternators and voltage regulators
61	Alternator charge indicator
B+	Battery positive terminal
B–	Battery negative terminal
D+	Alternator positive terminal
D–	Alternator negative terminal
DF	Alternator field winding
DF1	Alternator field winding 1
DF2	Alternator field winding 2
U, V, W	Three-phase terminals
	Audio systems
75	Radio, cigarette lighter
76	Loudspeaker
	Switches
	NC contact/changeover contact
81	Input
81a	1st output, NC side
81b	2nd output, NC side
	NO contact
82	Input
82a	1st output
82b	2nd output
82z	1st input
82y	2nd input
	Multi-stage switch
83	Input
83a	Output, position 1
83b	Output, position 2
83L	Output, position left
83R	Output, position right

Terminal	Definition
	Current relay
84	Input, output, relay contact
84a	Output, drive
84b	Output, relay contact
	Switching relay
85	Output, drive (end of winding negative or ground)
86	Input, drive (start of winding)
86a	Start of winding / 1st winding
86b	Winding tap / 2nd winding
	Relay contact for NC contact and changeover contact:
87	Input
87a	1st output (NC side)
87b	2nd output
87c	3rd output
87z	1st input
87y	2nd input
87x	3rd input
	Relay contact for NO contact:
88	Input
	Relay contact for NO contact and changeover contact (NO contact side):
88a	1st output
88b	2nd output
88c	3rd output
	Relay contact for NO contact:
88z	1st input
88y	2nd input
88x	3rd input
	Turn-signal lamp (turn-signal flasher)
C	1st indicator light
C0	Main terminal for check circuits separate from flasher
C2	2nd indicator lamp
C3	3rd indicator lamp (e.g for dual trailer operation)
L	Left-side turn-signal lamp
R	Right-side turn-signal lamp

References

[1] EN 60617: Graphical symbols for diagrams.
Part 2: Symbol elements, qualifying symbols and other symbols having general application.
Part 3: Conductors and connecting devices.
Part 4: Basic passive components.
Part 5: Semiconductors and electron tubes.
Part 6: Production and conversion of electrical energy.
Part 7: Switchgear, controlgear and protective devices.
Part 8: Measuring instruments, lamps and signalling devices.
Part 9: Telecommunications: Switching and peripheral equipment.
Part 10: Telecommunications: Transmission.
Part 11: Architectural and topographical installation plans and diagrams.
Part 12: Binary logic elements.
Part 13: Analogue elements.

[2] EN 81346: Industrial systems, installations and equipment and industrial products – Structuring principles and reference designations.
Part 1: Basic rules.
Part 2: Classification of objects and codes for classes.

[3] DIN 72552: Terminal markings for motor vehicles;
Part 1: Scope, principles, requirements.
Part 2: Codes.
Part 3: Examples for application on circuit diagrams.
Part 4: Summary.

Vehicle electrical systems for hybrid and electric vehicles

While the electrical system of a vehicle with start/stop system (micro hybrid) is very similar to a conventional vehicle electrical system (see 12 V vehicle electrical system), electrical systems for mild- and full-hybrid vehicles and externally rechargeable hybrid vehicles and electric vehicles usually have a high voltage level. They therefore differ significantly from the electrical system of a conventional vehicle.

The electrical system of a hybrid or electric vehicle has the following chief functions:

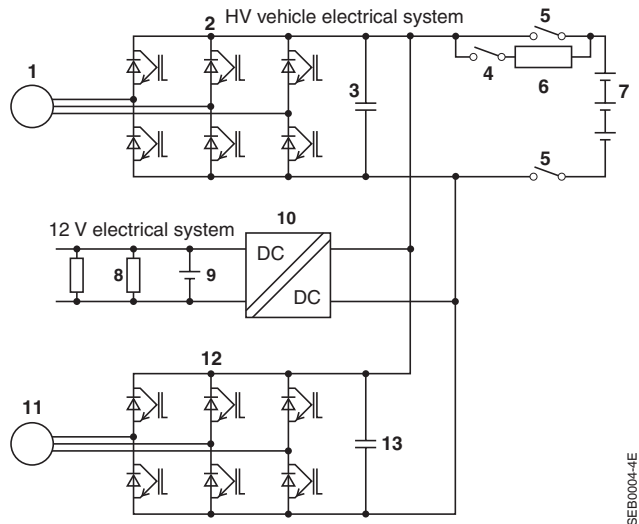
– storing excess electrical energy from the drivetrain,
– where necessary outputting electrical energy to the drivetrain,
– reliable supply of the electrical consumers,
– for externally rechargeable hybrid and electric vehicles, storing energy from the public power-supply system and supplying the vehicle with it during operation.

Vehicle electrical systems for mild- and full-hybrid vehicles

The function of a mild- or full-hybrid vehicle (see Hybrid drives) requires a large amount of electric power, 15 to 100 kW. This cannot be provided on the 12 V voltage level in a sensible way. Therefore an HV vehicle electrical system (HV stands for High Voltage) with a voltage in the range of 42 to 800 V is additionally required. However, the 12 V standard electrical system is still required to supply the 12 V consumers in the vehicle, especially the ECUs. The individual consumers are supplied from the corresponding vehicle electrical system according to their power requirements (Figure 1). Generally, for cost reasons, automakers attempt to make do with standard 12 V components, as these are available cheaply and in large quantities.

Figure 1: Electrical system of a parallel-hybrid vehicle (typically only one inverter is used for the drive)

1 Electrical machine,
2 PWM inverter,
3 DC link capacitor,
4 Pre-charging contactor,
5 Main contactor,
6 Pre-charging resistor,
7 HV battery,
8 12 V consumer,
9 12 V battery,
10 Isolating DC-DC converter,
11 Air-conditioning compressor,
12 PWM inverter,
13 DC link capacitor.



HV vehicle electrical system

Design

The HV vehicle electrical system consists of an HV battery, at least one PWM inverter to power the electric drive (electrical machine, e-motor), other high-output or HV consumers (e.g., electric air-conditioning compressor) and for voltage-safety reasons an isolating DC-DC converter to supply the 12 V vehicle electrical system (Figure 1).

Operating principle

The HV vehicle electrical system is supplied with voltage via the contactors integrated in the HV battery. When the vehicle is switched off or if there is an accident, the HV vehicle electrical system is de-energized via the contactors, limiting the hazardous high voltage to the battery pack.

The PWM inverter generates a three-phase alternating current from the high direct voltage with a variable current intensity and frequency for the electrical machine (see Inverters).

Power is supplied to the HV vehicle electrical system in hybrid drives when the electrical machine is in generator mode or from the HV battery.

The isolating DC-DC converter transfers electrical energy from the more powerful HV vehicle electrical system (typically 2 to 3 kW power output) to the 12 V vehicle electrical system. Power is thus supplied to this low-voltage vehicle electrical system from the HV vehicle electrical system.

An example of an additional component of the HV vehicle electrical system can be an electric air-conditioning compressor, which depending on the vehicle requires a maximum electric power of 3 to 6 kW and is activated by a second PWM inverter.

HV vehicle electrical system for parallel hybrids

The topologies of the vehicle electrical systems for mild and full hybrids with only one electrical machine each are very similar. The vehicle electrical system of a mild hybrid (compared to the full hybrid) makes do with a lower energy accumulation and lower output capacity, as the vehicle can "crawl" electrically for a very short time at most.

The primary flow of energy of a vehicle with parallel-hybrid drive is from the HV battery to the electrical machine and vice versa. There is also a small flow of energy via the DC-DC converter to the 12 V vehicle electrical system or to other HV consumers.

HV electrical system for power-split or series hybrid drives

Vehicles with two electrical machines require an enhanced electrical-system topology.

In the case of "Dedicated Hybrid Transmissions" (DHTs), i.e., power-split or parallel-series drivetrain topologies, high amounts of energy are transferred by the generator machine that can be coupled to the internal-combustion engine to the drive machine. Utilization of the optimal, i.e., low-loss, map ranges must be ensured in the operating strategy.

The battery voltage is determined by their design (cell number and size) according to the required temporary storage of electrical energy (typically for example for full hybrids 200 to 400 V). For optimum-efficiency operation of the electrical machine the DC link voltage of the inverters can however by variably controlled between the battery voltage and a maximum value set by the semiconductors (typically 600 to 700 V). Voltage adaptation and energy transfer are carried out by a suitably powerful step-up converter (non-isolating DC-DC converter, see Figure 2). By managing the DC link voltage just above the maximum value of the rectified induced voltages of the electrical machines it is possible to minimize the switching rate of the inverter switches and with it the electrical inverter losses and additional losses in the machines.

Further HV consumers like the air-conditioning compressor and air or water heater are usually supplied directly from the battery voltage.

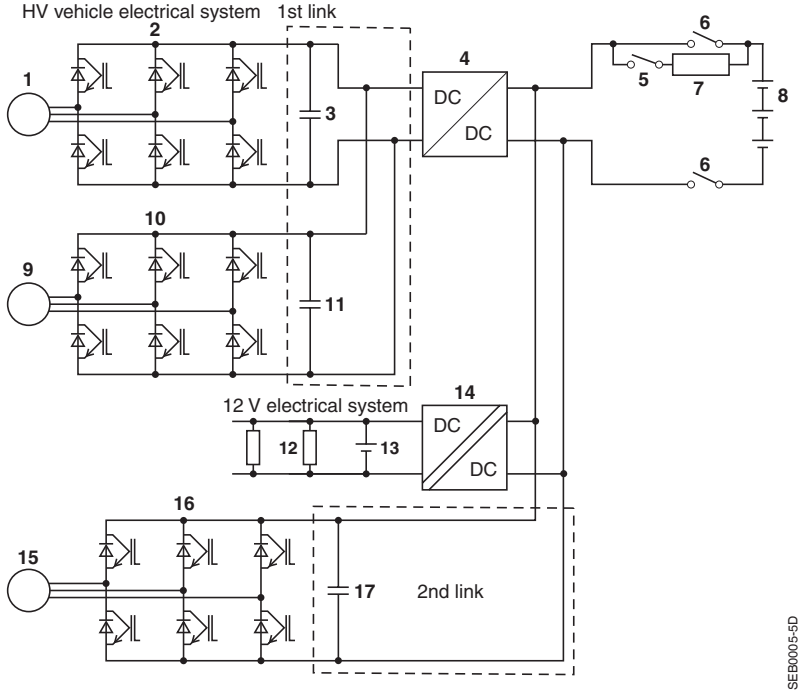
12 V vehicle electrical system

The 12 V vehicle electrical system has a similar design for all hybrid vehicles that have both HV and 12 V vehicle electrical systems. It is very similar to the 12 V electrical system of a conventionally powered vehicle except for the fact that it usually does not have a starter and the power is supplied not by an alternator, but by an isolating DC-DC converter from the HV machine and thus from the HV vehicle electrical system (Figure 2).

If the vehicle can drive or crawl electrically, all assisting functions – such as power steering, cooler pumps or braking system – must be operated independently of the internal-combustion engine, i.e., electrically, so that they are also available when the internal-combustion engine is switched off.

Figure 2: Electrical system of a power-split or series hybrid vehicle

1 Electrical machine, 2 PWM inverter, 3 DC link capacitor,
4 Non-isolating high-output DC-DC converter (step-up converter),
5 Pre-charging contactor, 6 Main contactor, 7 Pre-charging resistor,
8 HV battery, 9 Electrical machine, 10 PWM inverter,
11 DC link capacitor, 12 12 V consumer, 13 12 V battery,
14 Isolating DC-DC converter, 15 Air-conditioning compressor,
16 PWM inverter, 17 DC link capacitor.



Vehicle electrical systems for externally rechargeable hybrid and electric vehicles

Topology

The electrical system of an externally rechargeable hybrid vehicle (e.g., plug-in hybrid) mostly corresponds – other than having a larger battery with a greater energy-storage capacity and a charging device – to that of the comparable hybrid vehicle without external charging capability.

For a purely electric vehicle, the topology of the parallel-hybrid vehicle with a larger battery and a charging device is used in the simplest case (Figure 3). A charging device can be used that draws electrical energy from a public alternating-current power-supply system (230 V~) or a three-phase-current power-supply system (400 V~) and uses it to charge the battery in a controlled manner. The vehicle's battery management system tells the charging device the charging output it needs. For safety and to protect against overvoltage, this charging device usually has an electrically isolated design.

Alternatively, the battery of an electric vehicle can also be charged with direct current from an externally controlled charging station. In this case, all that is required is a plug-in connection to the HV battery and a communication interface between the battery management and charging station to specify the charging output.

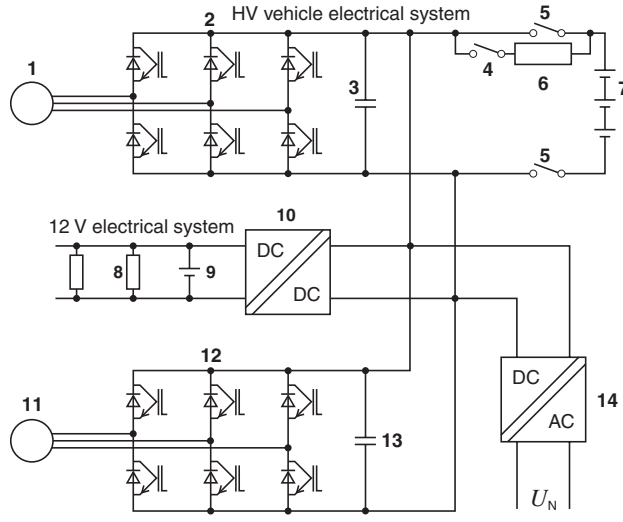
Efficiency

For electric vehicles, the efficiency of the vehicle electrical system is critical, as it has a direct effect not only on consumption, but also the range and the battery size needed. As a result, more expensive and more efficient components can reduce the total costs of the vehicle, as they make the same range possible with a smaller battery. Therefore, for many components of the vehicle electrical system, a review should be carried out to determine whether a less-expensive standard component can be used or it is worthwhile to develop a new, more efficient component.

Particular importance is attached to the energy demand for passenger-compartment climate control.

Figure 3: Electrical system of an electric vehicle

1 Electrical machine,
2 PWM inverter,
3 DC link capacitor,
4 Pre-charging contactor,
5 Main contactor,
6 Pre-charging resistor,
7 HV battery,
8 12 V consumer,
9 12 V battery,
10 Isolating DC-DC converter,
11 Air-conditioning compressor,
12 PWM inverter,
13 DC link capacitor,
14 Charging device (AC/DC converter).



Charging strategy

Cyclical variation
Generally, a battery is damaged and therefore its remaining service life shortened by cyclical charging and discharging ("cyclical variation"). This damage increases exponentially with the magnitude of the stored and re-drawn energy packages and thus the temperature variations of the cells. This cyclical variation is however indispensable for operation of the drivetrain for electric driving and braking (recuperation). The battery is designed on the basis of power output and range and the charging strategy is adapted within the context of optimization between service life, cost, and weight of the battery on the one hand and good hybrid-drivetrain efficiency and greater range for electric vehicles on the other hand.

Usually, the thermal aging processes of the battery cells are accelerated significantly at a high state of charge (SOC). Therefore, the charging strategy should avoid the combination of a high state of charge and high temperature of the cells.

Operating strategy for hybrid vehicles
Normally, the system attempts to keep the battery of a hybrid vehicle in an SOC window of approx. 50 to 70 %. If this window is exceeded, there can be no operating-point shift of the internal-combustion engine and no more recuperation.

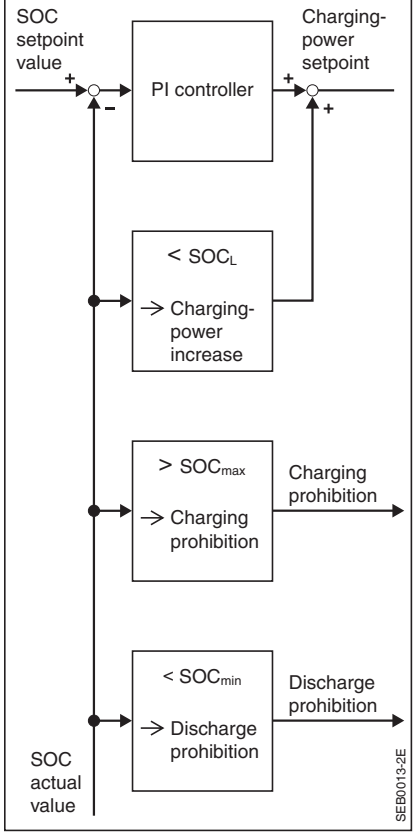
When the lower SOC limit (SOC_L) of approx. 50 % is reached, it is necessary to ensure that an adequate battery discharge power is still possible. Therefore, when the lower SOC limit is reached, battery recharging increases if no boost energy is required at the moment. Only when a much lower SOC limit is reached is the discharge power slowly reduced to zero. In normal vehicle operation, the vehicle almost never reaches this lower discharge limit, and the driver always experiences nearly identical acceleration behavior.

The lower SOC limit also plays an important role for reliably maintaining startability and preventing exhaustive discharge, which is detrimental to service life.

The charging strategy (Figure 4) is usually implemented in the engine control unit or in a special hybrid control unit.

Operating strategy for electric vehicles
In a pure electric vehicle, there is only one source of energy – the battery. Therefore, the operating strategy must fulfill the driver's command from this source, regardless of the state of charge. Therefore, strong but less frequent cyclical variation is inevitable to attain the necessary range of the vehicle with a reasonable battery size. The SOC limits are set here in such a way that there no effects that compromise service life. Depending on the battery system, the usable SOC range is between approx. 10 % and 90 %.

Figure 4: Charging strategy of a hybrid vehicle in driving mode
SOC State of Charge,
SOC_{max} Upper SOC limit,
SOC_L Lower SOC limit,
SOC_{min} Lower discharge limit.



Operating strategy for electric vehicles with range extender

For an electric vehicle with range extender – a small internal-combustion engine with an alternator that can supply electrical energy to the HV vehicle electrical system – there is the possibility during a longer journey or if a lower SOC limit value is reached of recharging the battery during the actual journey from this alternator instead of from an external source.

The range extender only has moderate rated power and delivers this as constant power with the best possible fuel consumption to the battery, which buffers the drivetrain power dynamics. There is also the possibility, subject to the appropriate drivetrain design, of coupling the internal-combustion engine of the range extender in a limited speed range directly to the drivetrain and thereby delivering mechanical power to the wheels without a loss of electrical efficiency.

Operating strategy for plug-in hybrid

The HV charging strategy here is a mix of that of an electric vehicle and that of a hybrid vehicle. In the case of a plug-in hybrid the aim for fuel-economy reasons is to drive electrically as far and as frequently as possible. When the battery is exhausted, the vehicle switches to an operating strategy similar to a hybrid (but with a lower SOC limit). However, the vehicle can also drive using only the internal-combustion engine where necessary.

When the battery is charged from the public power-supply system (typically in Germany 3.3 kW for single-phase current, in other countries up to 7.5 kW, up to 11 kW for three-phase current), the charging devices or the power output of the public power-supply system are usually limiting. The first phase of charging is performed at full power; in the second phase current and power are gradually decreased until the battery end-of-charge voltage is reached.

References

[1] S. Sasaki, E. Sato, M. Okamura: The Motor Control Technologies for the Hybrid Electric Vehicle; Toyota Motor Corporation.

Electrical machines for motor-vehicle propulsion

The electric drive for vehicles with a hybrid or a purely electric drive is typically supplied by an HV battery (see Batteries for electric and hybrid drives) with direct voltage, which an inverter converts into a symmetrical, three-phase alternating-voltage system (Figure 1). The voltage amplitude and the frequency of this alternating-current system can be changed. Figure 2 shows simplified characteristic torque, power, current and voltage curves obtained during inverter operation. In the base-speed range (between standstill and specific speed at which the maximum voltage is achieved in the alternating-voltage system) a constant torque is typically obtained. The range from the specific speed up to the maximum speed is called the field-weakening range. In this range the mechanical power of the drive is approximately constant, which corresponds to a torque reduction proportional to the reciprocal of the speed. However, the behavior in the field-weakening range is actually dependent on the machine type (see section Typical characteristic curves).

This review provides an overview of the most important features and influencing factors of electrical machines for motor-vehicle propulsion (see also [1] and [2]).

Requirements

Figure 3 shows by way of example important operating points for purely electric vehicle operation and their importance to driveability. From these the contents of three of the four requirement groups listed in Figure 4 can be derived: overload operation, continuous operation, and cycle efficiency. These groups and four others (oscillation excitations, installation space, strength, and costs) are explained in more detail in the following.

Figure 2: Simplified characteristic curves of an electric drive
1 Torque, 2 Mechanical power,
3 Line-to-line voltage U_L (fundamental component),
4 Supply current I (fundamental component).
A Base-speed range,
B Field-weakening range.

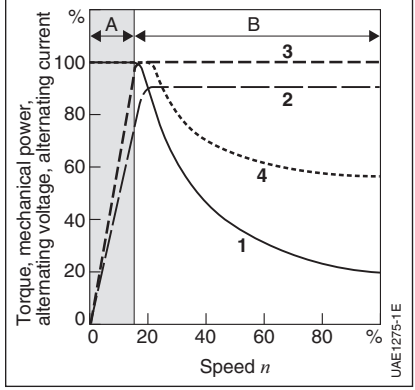
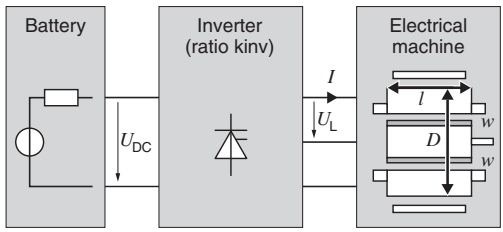


Figure 1: Electric drive
U_{DC} Battery voltage,
U_L Line-to-line voltage,
I Supply current from inverter to electrical machine,
l Axial length of laminated core of electrical machine,
D Outside diameter of stator,
w Number of turns.



Overload operation

Figure 5 shows typical characteristic torque and power curves of an electric vehicle drive. Overload operation characterizes driveability in accelerating and passing (overtaking) maneuvers. The electrical machine cannot be continuously operated in overload operation as this would result in at least one limit temperature (for example, winding or magnet temperature) being exceeded. Typical requirement variables are the maximum torque at low speeds, the power at high speeds, and an associated minimum operating time. The characteristic curves of maximum torque and maximum power against speed are independent of the thermal properties of the machine and are called maximum limit characteristic curves. These are determined by the electromagnetic properties of the machine, the maximum supply current, and the battery voltage.

Continuous operation

Continuous operation characterizes driveability during longer trips at approximately constant driving speed and is a measure of the thermal properties of the machine.

Figure 3: Important operating points for an electric motor-vehicle drive

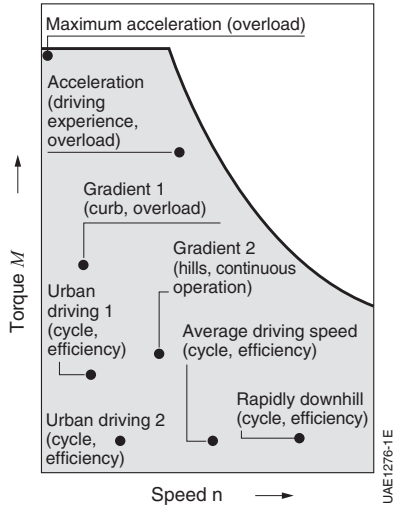


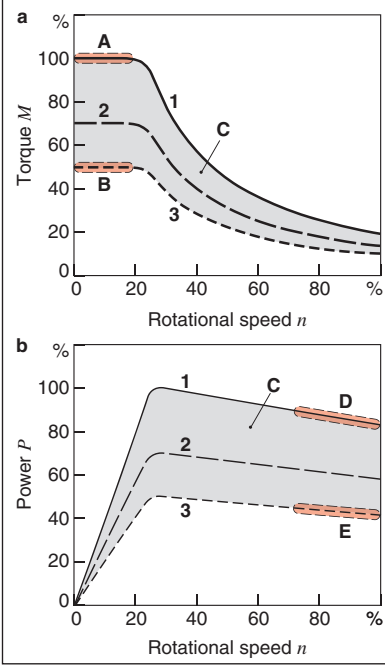
Figure 4: Requirement groups for an electric motor-vehicle drive

Machine design	Overload operation
	Continuous operation
	Cycle efficiency
	Oscillation excitations
	Installation space
	Strength
	Costs

UAE1277E

Figure 5: Typical characteristic torque and power curves for an electric motor-vehicle drive

a) Torque,
b) Power.
1 Maximum limit characteristic curve,
2 60 s characteristic curve,
3 Continuous limit characteristic curve.
A Maximum torque,
B Continuous torque at low speeds,
C Overload operation,
D Maximum power at high speeds,
E Continuous power at high speeds.



Continuous operation is defined by the reaching of a component limit temperature in thermal equilibrium. Typical requirement variables are the continuous torque at low speeds and a continuous power at high speeds. The characteristic curves of maximum continuous torque and maximum continuous power against speed are called continuous limit characteristic curves. Further characteristic curves with limited periods of time can be defined between the maximal limit characteristic curve and the continuous limit characteristic curve. Reaching the required period of time is dependent on the respective initial temperature.

Efficiency

Figure 6 shows the efficiency map of an electrical machine and the operating-point frequencies in a driving cycle. Depending on the distribution of the operating points in the cycle, other ranges are relevant to the cycle efficiency (typically rather small torques and medium speeds). Within the framework of design operating points below the maximum torque are as a rule calculated in such a way that optimum efficiency is obtained (regulation by the inverter, alternatively regulation to minimized oscillation excitations is also possible).

Oscillation excitations

In order to avoid unwanted sound radiation and service-life-reducing oscillations, oscillation excitations in the radial and tangential directions are calculated and minimized. The effects of torque ripple in the tangential direction depend on the mechanical systems connected to the machine (e.g., transmission). The requisite detailed knowledge of this system is often not available during the design process, which is why for the most part only a maximum peak-peak value of the torque-time characteristic is defined in the requirements.

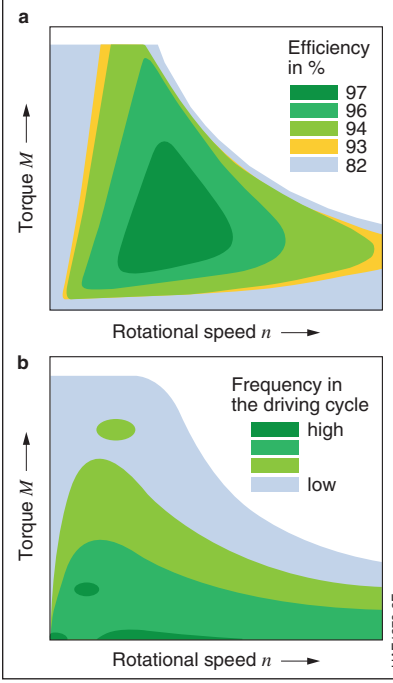
In the specifications for electrical machines as drives there are frequently maximum curves for the permissible sound-pressure or sound-power level. These can, depending on the quality of the mechanical model be calculated or estimated during the design process with the oscillation excitations in the radial direction.

Installation space

Especially for hybrid vehicles, which are still equipped with an internal-combustion engine as well as the electric drive, it is particularly important to minimize the machine size due to the small installation space available. The smaller this installation space, the greater the utilization of the electromagnetically active parts (laminated cores and windings). For machines which are integrated in the drivetrain of an internal-combustion engine (IMG, Integrated Motor-Generator), it may be sensible for example to prefer the electromagnetically less favorable individual-tooth winding of the distributed winding because the former saves a huge amount of installation space thanks to its very short winding overhangs in the axial direction.

Figure 6: Efficiency map and operating-point frequencies for an electric motor-vehicle drive

a) Efficiency map,
b) Operating-point frequency in the driving cycle.



Strength
The electromagnetic-design requirements frequently conflict with the mechanical-design requirements. Narrow bars are often necessary in the core laminations to direct the torque-generating magnetic flux in the desired direction. These restriction can however become critical from a mechanical perspective if they are no longer able to withstand the loads that occur at high rotational speeds. A coupled optimization, which optimizes the combination of the two functions, is necessary here.

Costs
The development of electrical machines as a motor-vehicle drive at Bosch is geared towards projects with high quantities (> 100,000 units per year). In series production in this field even seemingly small cost savings per machine noticeably pay off overall. Because of the large number of competitors the target market price can only be achieved through massive cost optimization. The foundation for this is laid by a volume-production design with as little materials usage as possible. The rare-earth magnets in permanent-field synchronous machines are, on account of the high prices of the raw materials neodymium and dysprosium, a particularly important cost factor.

From the vehicle manufacturer's standpoint the system costs of the overall drive are to the fore (sum total of battery, inverter, and electrical machine). In the system consideration it is sensible to optimize the components not individually and independently, but taking the overall system into account. This produces systems whose individual components are not necessarily optimum in terms of costs, but are so in terms of the overall system.

Design

Figures 7 and 8 show exploded view of the two most important types of electrical machine for motor-vehicle propulsion. Both machine types are permanent-field synchronous machines (PSM) with cuboidal neodymium-iron-boron magnet blocks which due to centrifugal force load are inserted as "buried magnets" in recesses in the laminated rotor core.

Integrated motor-generator
An integrated motor-generator (IMG) is located in a hybrid vehicle with the internal-combustion engine directly in the drivetrain. It maximum rotational speed is therefore the same as that of the internal-combustion engine. In view of the limited installation space in the axial direction and the high torque required to start the internal-combustion engine smoothly, the laminated cores of this machine type are usually axially short, but have a relative large diameter (typically around 300 mm).

Figure 7 shows the exploded view of an IMG. The stator in this case consists of single tooth-wound coils. These are first wound individually and then pressed into the stator housing. The single tooth-wound coils are contacted via the contact carrier. To this end the wire ends of the single tooth-wound coils are bent onto and connected with the respective contacting rail. The HV connection is joined to the contact carrier. The completed stator is secured for example by means of lugs in the transmission housing of the internal-combustion engine. The inside diameter of the transmission housing is slightly larger than the outside diameter of the stator housing in such a way that a gap is formed between the two. This gap is sealed with the aid of seals (not pictured here) and forms the cooling jacket for liquid cooling.

The heart of the rotor is the rotor carrier. This is usually hollow on account of the very limited installation space. Further components, such as for example a clutch, the torsional-vibration damper or hydrodynamic converters, are mounted inside the rotor carrier.

The rotor laminations with the permanent magnets are pressed onto the rotor

carrier and secured with a support ring. By turning the individual support rings against each other, it is possible to establish a stepping or skewing in the rotor which has positive effects with regard to noise generation in the machine. The rotor laminations can be provided with recesses which, in combination with openings in the rotor carrier and a wet-running clutch, provide spray cooling for the winding overhangs of the single tooth-wound coils. The resolver wheel is also mounted on the rotor carrier and serves in later assembly in combination with the resolver coils not pictured here to determine the position and speed of the rotor. The rotor

Figure 7: Design of an integrated motor-generator (IMG)
1 Stator housing, 2 Contact carrier, 3 Single tooth-wound coil, 4 HV connection,
5 Rotor carrier, 6 Rotor laminations, 7 Resolver wheel, 8 Permanent magnet, 9 Support ring.

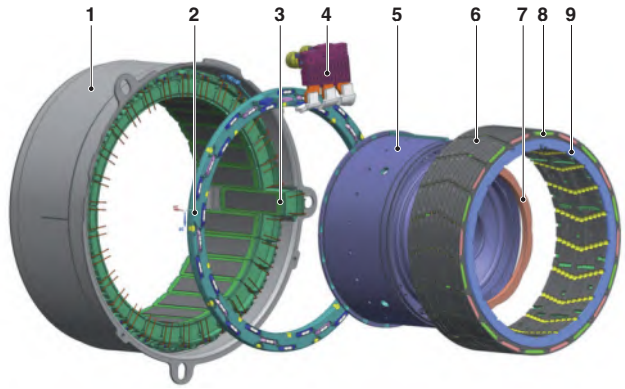
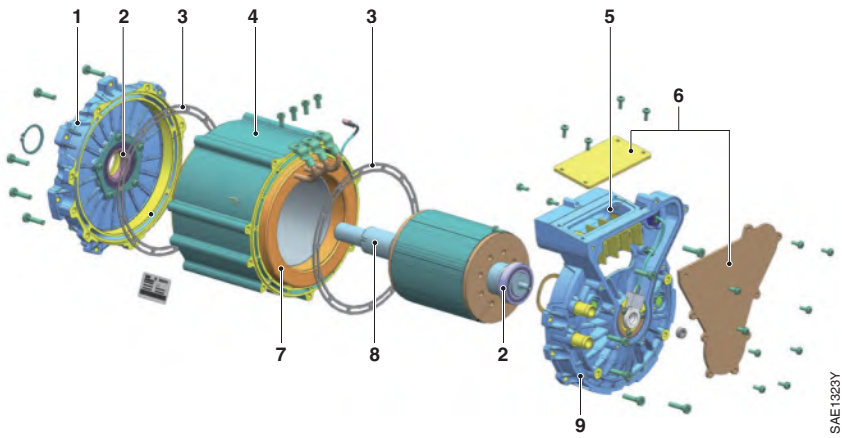


Figure 8: Design of a separate motor-generator (SMG)
1 End shield, A-side (drive-end), 2 Rotor bearings, 3 Seal, 4 Housing with water jacket,
5 Integrated connection box, 6 Covers for connection box,
7 Laminated stator core with stator winding, 8 Shaft end, A-side, 9 End shield, B-side.



carrier itself is connected with the transmission when the drivetrain is assembled and thereby ensures the transmission of torque.

Separate motor-generator
A separate motor-generator (SMG) can be connected with the drivetrain in different ways. This machine type is often used as a final-drive unit with a fixed-ratio transmission in hybrid or electric vehicles. Further applications include the installation of active parts in an electrified transmission or in an "electric axle" which forms a complete unit comprising inverter, transmission, and electrical machine (see eAxle).

Figure 8 shows the exploded view of a typical SMG as a "stand-alone machine" with water-jacket cooling. The drive-end end shield with the shaft end (A-side) is situated on the left. The connection of the water jacket in the housing with the end-face deflection of the coolant in the end shield is provided with seal. The laminated stator core with the stator winding is accommodated inside the housing. Because it is better suited to highly automated production, this machine type is frequently designed with a plug-in winding, where insulated copper rods are inserted into the stator slots, and the ends are bent to other rod ends and then welded together. On the B-side there is another end shield; this time with an integrated connection box for the power supply to the winding. The connection box is sealed with two covers so that the supply lines for the voltage source (inverter) can be installed. The rotor bearings are located in the end shields.

Types of electrical machines

Table 1 provides an overview of the most important types of electrical machines used in hybrid and electric vehicles. In hybrid vehicles a small installation space is for the most part decisive. This requirement can best be fulfilled with a permanent-field synchronous machine (PSM). The biggest drawbacks of this machine type are the relatively expensive rare-earth magnets in the rotor and the danger of these magnets being demagnetized.

Today's pure electric vehicles also contain induction or asynchronous machines (ASM) and electrically excited synchronous machines (ESM). Asynchronous machines with squirrel-cage rotors are characterized by simple, robust construction, but often require rotor cooling due to the additionally occurring rotor heat losses due to current. A very interesting alternative is the electrically excited synchronous machine, which on account of controllable rotor magnetization achieves particularly high efficiency levels and can manage on particularly little supply current. These advantages can lead to lower system costs and reduced energy consumption.

The switched reluctance machine (SRM) is often examined as an alternative with a particularly simple rotor design, but has to date been unable to gain a foothold in any production vehicle due to the increased complexity of the power electronics (3H-bridge instead of B6-bridge, see [1]) and its inherent susceptibility to noise generation.

Typical characteristic curves
The four types of electrical machine in Table 1 essentially come into question for use as traction drives in electric and hybrid vehicles in view of general considerations. Ultimately, the classic machine types are favored: asynchronous machine (ASM), permanent-field synchronous machine (PSM), and electrically excited synchronous machine (ESM) since these best fulfill the functional and economical requirements in the field of electrification.

Which of these three machine types is the right solution in any concrete application case can only be ascertained by means of a detailed comparison in which

all three machines are designed according to the same requirements and the same installation-space and electrical boundary conditions.

This comparison was conducted on the example of an electric vehicle (220 kW). The result is shown in the form of maximum limit characteristic curves with referred axis quantities in Figure 9. The abscissa represents the rotational speed referred to the maximum machine speed, the ordinate the mechanical drive power referred to the maximum available electric power. The differences between the three machine types on account of the principles involved can be clearly recognized after the maximum mechanical power is reached.

Asynchronous machine
The power of the asynchronous machine (ASM) drops at a markedly greater rate in the direction of maximum speed. The reason for this is the leakage inductance in the stator and rotor, which significantly influences the operational performance in this range. Because of this performance

Figure 9: Mechanical power for different types of electrical machines
1 ESM: Electrically excited synchronous machine,
2 PSM: Permanent-field synchronous machine,
3 ASM: Asynchronous machine.

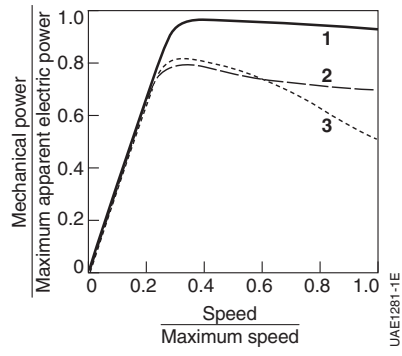
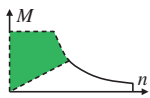
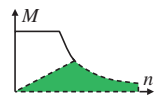
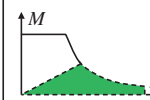
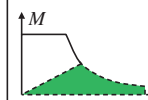


Table 1: Important machine types for electric motor-vehicle drives

Technology	**PSM** Permanent-field synchronous machine	**ESM** Electrically excited synchronous machine	**SRM** Switched reluctance machine	**ASM** Asynchronous machine
Functioning principle	Synchronous	Synchronous	Synchronous	Asynchronous
Torque density	Very high	Very high	High	High
Improvement in efficiency at	high torque low speed 	high speed low torque 	high speed low torque 	high speed low torque 
Potential	Minimum installation space No rotor cooling	High efficiency Low supply-current demand	Very simple rotor design	Simple rotor design
Rotor cooling	No/simple rotor cooling	Additional rotor cooling possibly required	No/simple rotor cooling	Additional rotor cooling possibly required
Risks	Magnet price Demagnetization Short-circuit resistance	Current-transformer system, rotor Speed strength	Noise Outlay, power electronics	High current demand Rotor heating

SAE1324-1Y

the ASM is less suited to drive cases which also require high power at maximum speed. This is basically the case in expressway/interstate driving close to the vehicle's top speed.

Permanent-field synchronous machine
The characteristic curve of the permanent-field synchronous machine (PSM) is much flatter. Here the amount of magnet material used (for the most part neodymium iron boron, NdFeB, due to the high flux density of these magnets) and the manifestation of the reluctance torque of the magnetic circuit offers plenty of scope in the layout of the characteristic curve. The PSM is therefore highly suitable for universal use and is the preferred solution for drives with high power and torque densities, provided the expensive rare-earth magnets permit economic feasibility of the drive.

Electrically excited synchronous machine
The electrically excited synchronous machine (ESM) has two particularities distinguishing it from the ASM and the PSM: Firstly, the maximum achievable mechanical drive power is for the most part clearly above the other two machine types and, secondly, the characteristic curve runs almost horizontally up to the maximum speed. This is possible due to the additional controllability of the excitation current in the rotor. The ESM therefore best converts the available electric power into mechanical power. Efficiency and power factor are accordingly higher for the ESM than for the ASM and the PSM. However, it is necessary to weigh up whether these advantages justify the design outlay for the excitation-current supply (collector rings or contactless transformers), and whether the effort and outlay associated with rotor cooling and mechanical strength at higher speeds justify its use as a driving machine.

Particularities of the machine types
Aside from the immediate properties which are derived from the characteristic curve, further particularities are to be noted in the machine types: In both synchronous machines requirements with regard to limiting the steady-state and dynamic short-circuit current dramatically influence the design specifically at high speeds. Small short-circuit currents necessitate particularly with the PSM low power at maximum speed.

In asynchronous machines much higher losses occur in the rotor. For this reason, active rotor cooling is for the most part necessary to achieve acceptable thermal conditions. Solutions with direct cooling of the short-circuiting rings in the rotor by oil or shaft cooling by water are possible here.

Essentially, the choice of the electrical machine suitable for the concrete drive case is the result of an overall evaluation of technical and economic characteristics which cannot generally or readily be transferred from one application case to the other. This statement also reflects the fact that all three favored electrical-machine types have been developed by different vehicle manufacturers to series maturity and are used in electrified drivetrains. A standard solution is not discernible at the present time.

Influencing factors on operational performance

The operational performance of the electric drive is influenced by the battery, the inverter, and the electrical machine. Important influencing factors on the maximum limit characteristic curve of the ASM and the PSM are presented in the following. Diagrams are only shown for the PSM, since the ASM and the PSM perform similarly in most cases.

Figure 1 shows the electric drive consisting of battery, inverter, and electrical machine, and visualizes the influencing variables considered
– battery voltage U_{DC},
– maximum supply current I_{max} of the inverter,
– axial length l of the laminated core,
– number of turns w,
– outside diameter D of the stator,
– turns ratio k of the inverter.

Battery voltage

A higher battery voltage increases the maximum line-to-line voltage U_L. This increases the specific speed of the drive (highest speed at maximum torque, Figure 10).

The battery voltage accordingly has no influence on the maximum torque in the base-speed range. The power in the field-weakening range on the other hand rises with the battery voltage. For the PSM the relationship is approximately linear and for the ASM approximately square.

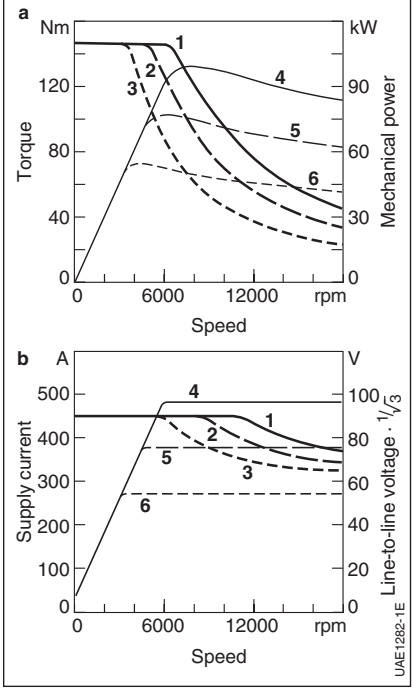
For the maximum limit characteristic curves in Figure 10 it is presupposed that the battery can also provide the necessary power for the maximum limit characteristic curve. Otherwise the achievable power on the maximal limit characteristic curve is limited by the battery power.

Maximum supply current of the inverter

The maximum supply current has no influence on the power at maximum speed, since the current limit of the inverter is usually not reached here (Figure 11). The maximum torque in the base-speed range on the other hand rises with the maximum supply current. Disregarding saturation, the relationship for the PSM is linear and for the ASM square. With a greater supply current the losses on the maximum limit characteristic curve also increase, above all in the base-speed range. This reduces the permissible overload time, i.e., the time until the electrical machine's temperature limit is reached.

Figure 10: Influence of battery voltage U_{DC} on the maximum limit characteristic curve of a PSM

a) Torque curve and mechanical power,
b) Supply current and line-to-line voltage.
1 Torque (a) and supply current (b) for U_{DC} = 230 V,
2 Torque (a) and supply current (b) for U_{DC} = 180 V,
3 Torque (a) and supply current (b) for U_{DC} = 130 V,
4 Mechanical power (a) and line-to-line voltage (b) for U_{DC} = 230 V,
5 Mechanical power (a) and line-to-line voltage (b) for U_{DC} = 180 V,
6 Mechanical power (a) and line-to-line voltage (b) for U_{DC} = 130 V.



Axial length of the electrical machine's laminated core

The maximum torque in the base-speed range is proportional to the axial length of the laminated core (Figure 12). For the PSM the power remains virtually constant at high speeds. For the ASM on the other hand the power decreases at high speeds with increasing axial length. The cause is the increased leakage inductance.

Figure 11: Influence of maximum supply current I_{max} on the maximum limit characteristic curve of a PSM

a) Torque curve and mechanical power,
b) Supply current and line-to-line voltage.
1 Torque (a) / supply current (b) for I_{max} = 520 A,
2 Torque (a) / supply current (b) for I_{max} = 450 A,
3 Torque (a) / supply current (b) for I_{max} = 380 A,
4 Mechanical power (a) / line-to-line voltage (b) for I_{max} = 520 A,
5 Mechanical power (a) / line-to-line voltage (b) for I_{max} = 450 A,
6 Mechanical power (a) / line-to-line voltage (b) for I_{max} = 380 A.

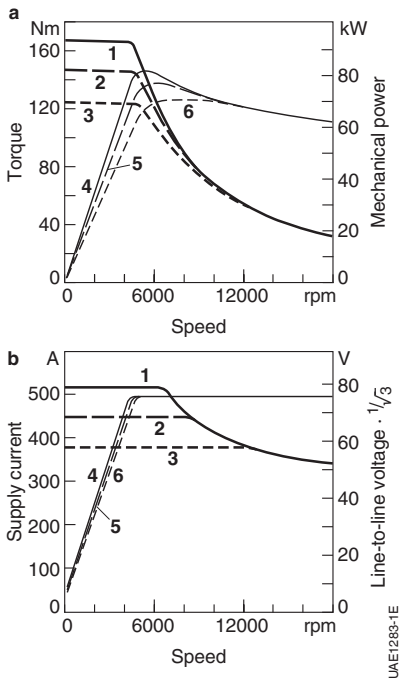
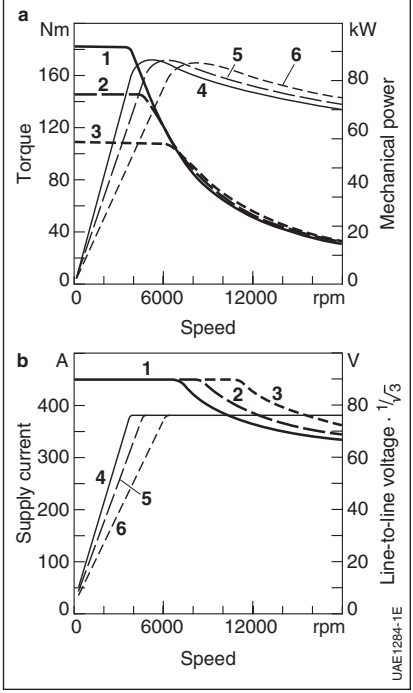


Figure 12: Influence of standardized axial length l/l_0 of the laminated core on the maximum limit characteristic curve of a PSM

a) Torque curve and mechanical power,
b) Supply current and line-to-line voltage.
1 Torque (a) / supply current (b) for l/l_0 = 125 %,
2 Torque (a) / supply current (b) for l/l_0 = 100 %,
3 Torque (a) / supply current (b) for l/l_0 = 75 %,
4 Mechanical power (a) / line-to-line voltage (b) for l/l_0 = 125 %,
5 Mechanical power (a) / line-to-line voltage (b) for l/l_0 = 100 %,
6 Mechanical power (a) / line-to-line voltage (b) for l/l_0 = 75 %.



Number of turns of the electrical machine

If the number of turns is increased on the ASM or the PSM, the maximum torque in the base-speed range increase on account of the higher slot current linkage (Figure 13). The power at high speeds is on the other hand reduced, since the inductance of the electric machine is proportional to the square of the number of turns and thus the maximum phase voltage is reached earlier. The slot's copper space factor (i.e., the ratio of copper space to slot space) is limited by the respective production engineering, therefore the conductor diameter must be re-

Figure 13: Influence of standardized number of turns w/w_0 on the maximum limit characteristic curve of a PSM

a) Torque curve and mechanical power,
b) Supply current and line-to-line voltage.
1 Torque (a) / supply current (b) for w/w_0 = 115 %,
2 Torque (a) / supply current (b) for w/w_0 = 100 %,
3 Torque (a) / supply current (b) for w/w_0 = 85 %,
4 Mechanical power (a) / line-to-line voltage (b) for w/w_0 = 115 %,
5 Mechanical power (a) / line-to-line voltage (b) for w/w_0 = 100 %,
6 Mechanical power (a) / line-to-line voltage (b) for w/w_0 = 85 %.

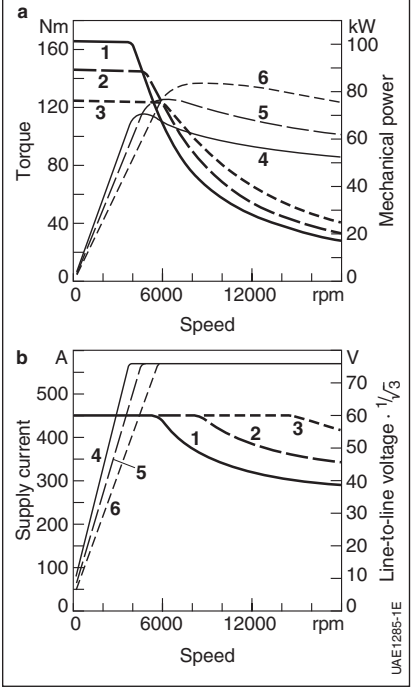
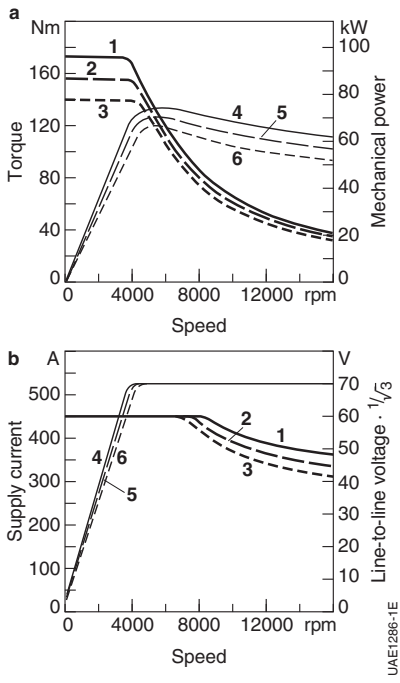


Figure 14: Influence of outside diameter D on the maximum limit characteristic curve of a PSM

a) Torque curve and mechanical power,
b) Supply current and line-to-line voltage.
1 Torque (a) / supply current (b) for D = 200 mm,
2 Torque (a) / supply current (b) for D = 180 mm,
3 Torque (a) / supply current (b) for D = 162 mm,
4 Mechanical power (a) / line-to-line voltage (b) for D = 200 mm,
5 Mechanical power (a) / line-to-line voltage (b) for D = 180 mm,
6 Mechanical power (a) / line-to-line voltage (b) for D = 162 mm.



duced if the number of turns is increased. In all, the winding resistance increases such that the losses on the maximum limit characteristic curve increase, above all in the base-speed range. This reduces the permissible overload time, i.e., the time until the electrical machine's temperature limit is reached.

Outside diameter of the stator

The influence of the stator outside diameter is considered whereby the entire lamination section of the electric machine is linearly scaled with the outside diameter. In the base-speed range this produces a greater torque when the outside diameter is increased for the ASM and the PSM (Figure 14). For the chosen scaling it must be noted that the maximum supply current of the inverter and the number of turns of the electrical machine have remained the same. With increasing outside diameter the current density decreases and the overload time increases.

Higher power is obtained for the PSM in the field-weakening range. For the ASM the power in the field-weakening range can also increase depending on design, but can also remain constant.

Turns ratio of the inverter

The turns ratio k describes the relationship between the battery voltage U_{DC} and the maximum voltage of the fundamental component $U_{L,max}$ (line-to-line voltage) of the alternating-current system:

$$U_{L,max} = \sqrt{3}\, k\, U_{DC}\,. \qquad (1)$$

Different modulation techniques which give rise to different turns ratios can be used in regulation of the inverter. Common techniques and the associated ideal turns ratios are

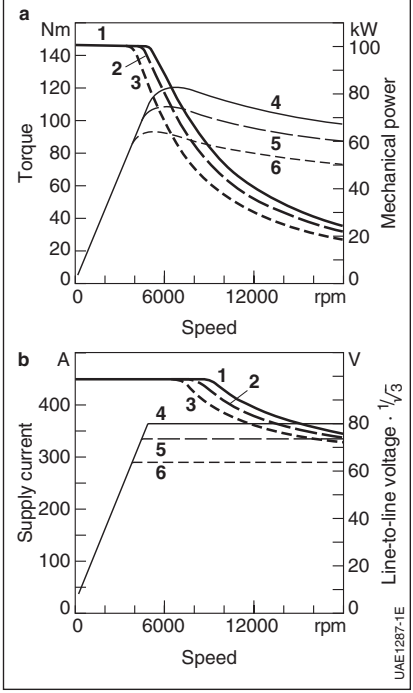
– sine-wave modulation, k = 0.3535,
– space vector modulation, SVM; space vector pulse width modulation, SVPWM, k = 0.4082,
– block clocking, k = 0.4502.

Voltage drops via the power semiconductors are disregarded in the specified values.

Depending on the modulation technique used, the battery voltage is utilized differently and there are different voltage limits (Figure 15). k therefore has no influence on the maximum torque in the base-speed range. With higher turns ratios the voltage limit is only reached at greater speeds so that higher power is achieved in the field-weakening range.

Figure 15: Influence of turns ratio k on the maximum limit characteristic curve of a PSM

a) Torque curve and mechanical power,
b) Supply current and line-to-line voltage.
1 Torque (a) / supply current (b) for k = 0.4502,
2 Torque (a) / supply current (b) for k = 0.4082,
3 Torque (a) / supply current (b) for k = 0.3535,
4 Mechanical power (a) / line-to-line voltage (b) for k = 0.4502,
5 Mechanical power (a) / line-to-line voltage (b) for k = 0.4082,
6 Mechanical power (a) / line-to-line voltage (b) for k = 0.3535.



Maximum apparent power
The maximum apparent power of the electric drive is

$$S_{max} = \sqrt{3}\ U_{L,max}\ I_{max}\ . \qquad (2)$$

For a given modulation technique the following thus applies

$$S_{max} \sim U_{DC}\ I_{max}\ . \qquad (3)$$

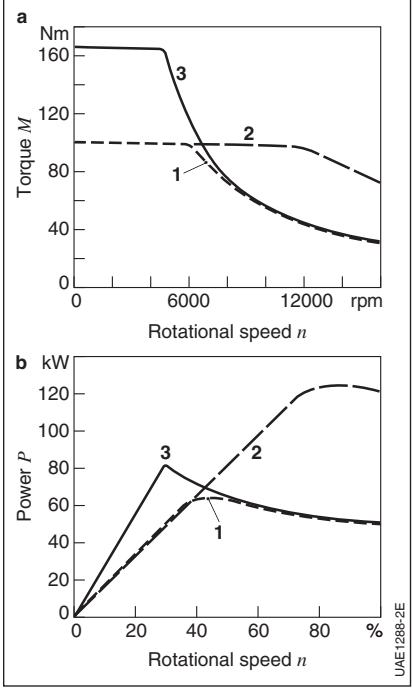
The maximum apparent power is decisive in determining which maximum limit characteristic curve can be achieved with an electric drive.

The behavior of the electric drive when the battery voltage (influence on the power at maximum speed), the maximum supply current (influence on the maximum torque in the base-speed range) and the number of turns of the electrical machine (exchange of torque in the base-speed range for power in the field-weakening range) are changed is described in the previous sections. In practice, however, these parameters are not adapted alone, but matched to each other by reference to the available maximum apparent power S_{max}.

By varying the number of turns the same maximum limit characteristic curve can be set for different combinations of U_{DC} and I_{max} as long as S_{max} remains constant. This is illustrated with reference to Figure 16. The starting position is an electric drive with the battery voltage $U_{DC} = U_0$, the maximum phase current $I_{max} = I_0$, and the number of turns $w = w_0$. This has the maximum limit characteristic curve 1 from Figure 16. If the battery voltage is now doubled, this produces the characteristic curve 2 from Figure 16 (similar to Figure 10); the maximum apparent power is doubled. To obtain the starting characteristic curve again, first the number of turns w must be doubled (characteristic curve 3 in Figure 16, similar to Figure 13) and then additionally the maximum phase current I_{max} halved (characteristic curve 1 in Figure 16, similar to Figure 11).

Adapting the number of turns produces in both cases (both settings for characteristic curve 1 in Figure 16) the same maximum limit characteristic curve. Battery voltage and maximum phase current are indeed different, but the maximum apparent power is according to equation (3) constant. It must be noted that the number of turns can only be changed in discrete steps and therefore deviations can arise in reality.

Figure 16: Setting the maximal limit characteristic curve of a PSM by varying the number of turns w at the same apparent power
Settings:
1 $U_{DC}/U_0 = 1$, $I_{max}/I_0 = 1$, $w/w_0 = 1$ and $U_{DC}/U_0 = 2$, $I_{max}/I_0 = 1/2$, $w/w_0 = 2$.
2 $U_{DC}/U_0 = 2$, $I_{max}/I_0 = 1$, $w/w_0 = 1$.
3 $U_{DC}/U_0 = 2$, $I_{max}/I_0 = 1$, $w/w_0 = 2$.



Losses in electrical machines

Figure 17 shows a schematic longitudinal section of an electrical machine in which the most important assemblies are indicated and the associated types of loss are assigned to these.

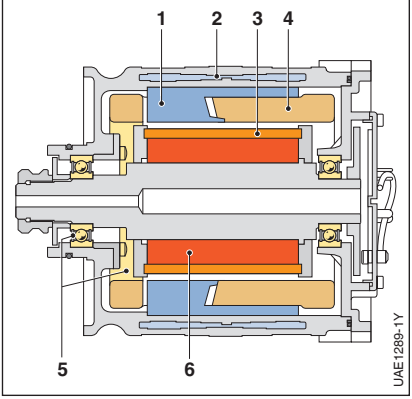
Heat losses due to current
Heat losses due to current which in the first approximation are proportional to the square of the effective value of the current occur in the stator winding, in the rotor case of an asynchronous machine, and in the excitation winding of an electrically excited synchronous machine. For optimized designs, current-displacement and proximity effects as well as parasitic circulating currents in parallel and delta connections must additionally be taken into account.

Iron losses
Iron losses occur due to rapid magnetic reversals predominantly in the laminated stator core of the machines. On account of the principles involved, synchronous and asynchronous machines exhibit absolutely no or only low-frequency sinusoidal components of flux through their basic fields in the rotor. The magnetic losses are usually broken down into two components:
– hysteresis losses $\sim B^2 f$,
– eddy-current losses $\sim B^2 f^2$.

Here B denotes the flux density and f its frequency. Because of the high speeds and the accompanying high frequencies in automobile drives, the task of minimizing the eddy-current losses in particular poses a huge challenge. This is usually achieved by accordingly optimizing the lamination section and by using particularly thin core laminations (usually ≤ 0.35 mm) which are insulated against each other and stacked in the axial direction to form the laminated core.

Parasitic effects such as for example harmonic fields or harmonic components generated by the inverter can generate significant eddy-current losses both in the stator and in the rotor which must be taken into consideration in the design process.

Figure 17: Types of loss in an electrical machine
1 Iron losses in the stator lamination,
2 Eddy-current losses in the housing,
3 Ohmic losses in the rotor cage (ASM), ohmic losses in the excitation winding (ESM), eddy-current losses in the magnets (PSM),
4 Ohmic losses in the stator winding,
5 Friction losses (bearing friction and windage),
6 Iron losses in the laminated rotor core.



Eddy-current losses
Further eddy-current losses can occur in the magnet blocks of a permanent-field synchronous machine and in structural parts (e.g., housings or end shields) due to the magnetic alternating fields generated by the machine. Of particular importance here are the eddy-current losses in the magnets, since these directly heat the magnetic blocks which are thermally for the most part poorly connected to the cooling system and consequently under high field load can lead to irreversible demagnetization.

Friction losses
The rotational movement of the rotor causes bearing-friction and windage losses, which on the one hand reduce efficiency and on the other hand can even result in unacceptable heating in the bearing itself. Particularly in hybrid vehicles, in which the electrical machine always rotates but is not always actively in use, the friction losses have a significant influence on cycle efficiency.

Figure 18: Typical cooling types in an electrical machine

a) Water cooling,
b) Oil-jacket cooling and internal oil cooling.
1 Oil-jacket cooling, 2 Internal oil cooling,
3 Stator, 4 Rotor, 5 Oil supply,
6 Coolant supply from cooling circuit,
7 Coolant discharge to cooling circuit.

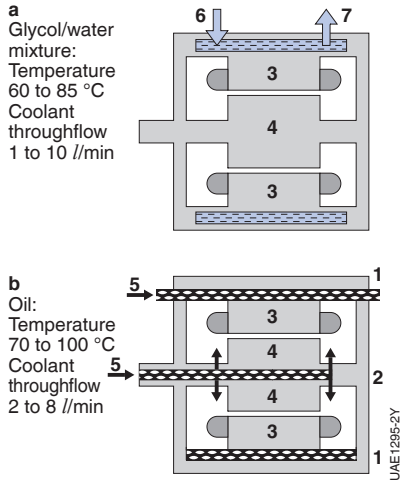
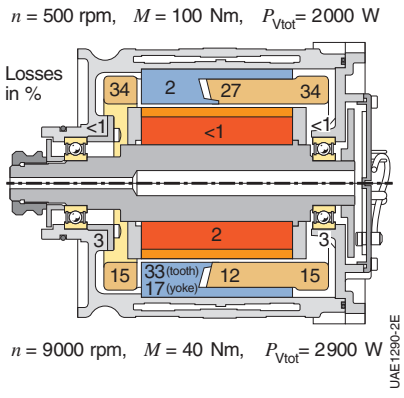


Figure 19: Typical distribution of losses of an electrical machine

Designations, see Figure 17. Percentage values in the top half apply to the operating state mentioned above. Values in the bottom half apply to the operating state mentioned below.



Cooling of the electrical machine

The losses are dissipated in the machine to cooling media. A powerful cooling system protects the individual components against overheating. Insufficient dissipation of the individual heat losses from the electrical machine restricts the permissible continuous power. In overload operation the thermal mass (heat capacity) limits the time until a limit temperature is reached.

Cooling types

Different cooling media are provided on the vehicle side to cool the electrical machine (Figure 18).

Water cooling
The type most commonly encountered is water cooling, i.e., integration in a water/glycol circuit, which can also be used to Cooling the power electronics. Here the stator core is encased in a water-cooling jacket.

Oil cooling
Oil cooling is advisable for integration in the manual transmission or final-drive differential. Cooling is also effected at the stator jacket. But the rotor and the stator winding can also be cooled directly.

Air cooling
Air cooling is also used. Here air flows through a fan via large-area cooling fins along the stator jacket or even axially along the rotor.

The cooling effect is crucially determined not just by the cooling type, but also by the coolant temperature and the coolant throughflow.

Heat dissipation

Evaluation of the cooling of an electrical machine requires knowledge of the individual, locally resolved losses and describes the heat paths to the cooling fluid. This is performed at the stator essentially by heat conduction from the winding to the cooling duct and at the rotor without internal cooling by convection. Mass node models and also CFD techniques (Com-

putational Fluid Dynamics) are used for design purposes.

In the case of jacket cooling the main stator heat flow (90 %) proceeds from the winding via the stator-slot insulating paper to the stator core and from there to the housing with cooling duct. A specific role is assigned to winding cooling here. Electric insulation by paint, impregnating resin and insulating paper causes heavily directionalized thermal insulation. Cooling optimization here requires the interaction of electromagnetic design, production development, and thermal design.

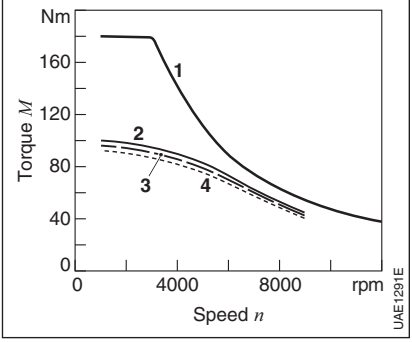
Heat dissipation via the stator winding overhangs to the cooling housing by air can be increased by fans or additional oil cooling. Alternatively, a heat-conducting connection by casting materials can be established.

The rotor is thermally connected via the air gap to the stator and convectively connected via the side faces to the housing. It is heated or cooled by the stator, depending on the operating situation.

The distribution of losses in the electrical machine changes with the speed (Figure 19). At low speed (top half of the picture) primarily heat losses due to current are effective; at high speed (bottom half of the picture) iron losses are predominant. In this way, the different cooling paths work on the basis of speed.

Figure 20: Example of a continuous limit characteristic curve of an electrical machine

Continuous limit characteristic curves with variation of the cooling-fluid temperature. Cooling-fluid throughflow is in all cases 6 l/min.
1 Maximum characteristic curve for torque,
2 Continuous torque for 65 °C,
3 Continuous torque for 75 °C,
4 Continuous torque for 85 °C.



Thermal limitation

When the electrical machine has a high power density, the maximum power is typically above the continuous power. Protection of the components against overheating is then ensured by a limiting function of the inverter's activation software. Limitation is performed typically for the limit temperatures of the stator and the rotor. In the case of the stator the insulation resistance limits the maximum winding temperature (e.g., 210 °C). Here the frequency distribution from vehicle operation is taken into account for the service-life evaluation. The rotor limit temperature is established for the PSM by the demagnetizing resistance of the magnets. This value must be derived from a worst-case analysis for the demagnetizing field at the limit temperature of the magnet (e.g., 160 °C).

Determination of the stator and magnet temperature during operation is subject to tolerances and requires a reduced control limit. Figure 20 shows by way of example the continuous limit characteristic curve of a water-cooled PSM. Limitation for the stator temperature is effective here. Lowering the coolant temperature increases the drivable continuous torque.

References

[1] K. Fuest, P. Döring: Elektrische Maschinen und Antriebe. 7th Ed., Vieweg + Teubner Verlag, 2007.

[2] A. Binder: Elektrische Maschinen und Antriebe. Springer-Verlag, 2012.

[3] R. Fischer: Elektrische Maschinen. 17th Ed., Carl Hanser Verlag, 2017.

Inverters for electric drives

Area of application

Function

Inverters – also known as converters – contain the power electronics used to supply multi-phase electrical machines (EM) (Figure 1). Exclusively PWM inverters are used for supplying direct voltage (direct current, DC) in automotive engineering. The direct voltage can, depending on the drive system, be supplied directly by the traction battery or via an upstream DC-DC converter for voltage adaptation.

Requirements

Compared with industrial inverters, inverters used in electric vehicles are subject to much more critical requirements or demands with regard to coolant temperatures, environmental factors (corrosion, vibration, and tightness demands), and in particular electrical and functional safety.

Currently the most challenging application for an inverter is its use in plug-in hybrids because here the installation spaces and the coolant temperatures are rendered particularly critical by the additional presence of an internal-combustion engine and a traction battery.

All the components in the HV circuit (High Voltage) of a vehicle (HV vehicle electrical system) are for touch-voltage safety reasons isolated from the vehicle body ("frame" or "ground" of the 12 V vehicle electrical system, body potential), the inverter housing, and the control circuits connected to the 12 V vehicle electrical system.

Performance classes

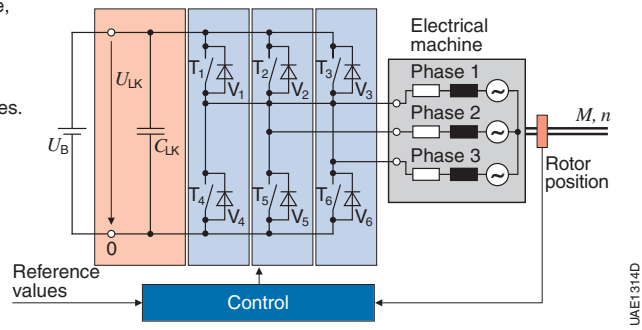
Electric drives are used in vehicles of widely differing performance classes. The drive power per drive in the passenger-car sector ranges between 60 and 200 kW for a battery end-of-charge voltage typically of 450 V. The three-phase machine currents ranges between 250 and 600 A. Two or three drives may be present in the vehicle, depending on the application.

Batteries with a voltage of 800 V are used for passenger cars in the upper performance segment and in the commercial-vehicle sector. The attainable power is therefore 250 to 600 kW at similar currents.

The current intensity is limited with the power-module technology used. Three-phase current systems with six phases can also be used with appropriately lower current intensity for high power.

Figure 1: Power circuit of a three-phase AC drive with inverter (block diagram with switches)

U_B Voltage of traction battery,
U_{ZK} DC link voltage,
C_{ZK} DC link capacitance,
M Motor torque,
n Motor speed.
T Power switches (power transistors),
V Free-wheeling diodes.



Function and circuitry

Three-phase current system

The inverter power switches "pulse" the direct voltage (DC) in specific switching patterns to the electrical-machine windings. Alternating voltages whose fundamental components generate the alternating currents (AC) in the electrical-machine stator windings are applied at the windings. In the case of polyphase (induction) machines there are at least three winding phases which are supplied with alternating currents that are symmetrically offset in relation to each other, the so-called three-phase current system.

Basic inverter circuit

Six power switches are required in an inverter's basic circuit. At each of the three winding phases one of the power switches is connected to the positive polarity and one is connected to the negative polarity of the feeding direct voltage (Figure 1). The power switches connected to the positive polarity are called "high-side switches" and those connected to the negative polarity are called "low-side switches". These six switches in all form a three-phase bridge connection. The combination of a high-side switch and a low-side switch is called a half-bridge. A three-phase bridge connection therefore consists of three half-bridges.

Design variations of power switches

Depending on the required current-carrying capacity, each of the power switches can consist of one or more parallel-connected individual transistors. IGBT transistors (Insulated Gate Bipolar Transistor) are used as power transistors in the vehicle traction-voltage range of 150 to 800 V. These transistors, however, only route the current in one direction so that antiparallel (in the direct-voltage reverse direction) connected diodes (free-wheeling diodes) are required for free-wheeling and recovery time ranges of the current characteristics. In recuperation mode these diodes direct the current generated by the winding phases to the battery.

It transpires that in the medium term power transistors made of SiC (silicon carbide) instead of pure Si are used in applications with higher supply voltages (e.g., 800 V) and to reduce losses. These have MOS-transistor properties, are consequently reverse-conducting and therefore do not require an antiparallel-connected diode.

Switching the machine currents

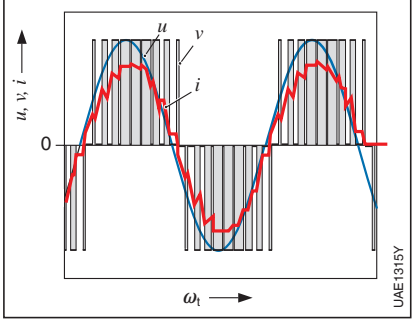
When the inverter is clocked, the voltages at the machine supply lines jump (phase voltages, Figure 2). In view of the inductances in the electrical machine, the alternating currents in the windows however have a strongly smoothed course, ideally a sinusoidal shape. The clock frequency T is for the most part constant, typically 10 kHz. The clock frequency can, depending on the operating point, be reduced in order to reduce switching losses in the inverter and additional losses in the electrical machine. This however results in a low-frequency noise excitation. The noise level may rise, depending on the natural frequencies of the mechanism.

The sinusoidal characteristic in Figure 2 illustrates the first order, i.e., the fundamental component from the Fourier analysis of the switched voltage characteristic. Harmonics are made visible in the current characteristic. These lead to harmonic fields and thus certain losses (additional losses) in the machine's iron core.

The sequential switching of the machine current between the two polarities

Figure 2: Time characteristics

v Phase voltage (measured to star point of machine winding),
u Sinusoidal fundamental component,
i Resulting machine-current characteristic (with harmonics),
ω_t Angular frequency.



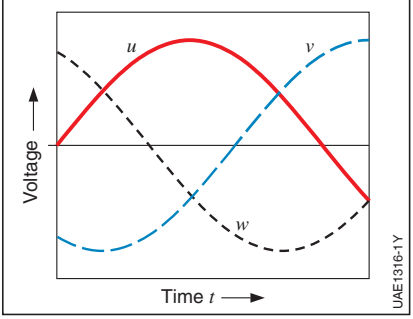
of the direct voltage (high- and low-sides) is called commutation. The switching patterns selected by the inverter control gives rise to an effective machine terminal voltage which drive the required instantaneous machine alternating current (Figure 3). The amplitude of the phase currents in the machine supply lines essentially determines the machine torque, the frequency of the current determines the rotational speed. The sum total of the voltage-time phases produces in conjunction with the phase currents a continuous active-power flow on the DC side of the inverter.

Function of the link

Parasitic inductances in the mechanical circuit arrangement create voltage peaks at the power switches during commutation. The link capacitor is required to reduce these voltage peaks and the AC component (ripple content) in the battery voltage and to smooth the power pulses in the direction of the traction battery. It must meet strict demands with regard to low-inductance design, low losses, and high current-carrying capacity. Film capacitors which offer a good compromise between performance and cost are currently used here. Power switches and link must be set up in a low-inductance arrangement so that the resulting commutation circuit can achieve rapid current build-up and current reduction in the transistors of the power switches and free-wheeling diodes with simultaneously low voltage peaks.

Figure 3: Ideal three-phase voltage system with time characteristic

u, *v*, *w* Phase voltages.



Secondary functions of the inverter

As well as supplying the electrical machine, the inverter has the following secondary functions:

– Discharging the direct-voltage link during deactivation and in an emergency.
– Protecting the traction electrical system against overvoltage.

Overvoltage protection

Overvoltage protection in the event of a fault with a permanent-field synchronous machine (PSM) as the result of its rotational-speed-induced voltage is particularly critical. In the event of an inverter-control failure, the induced machine voltages are rectified via the free-wheeling diodes of the PWR inverter. When the battery is connected a braking torque builds up at the machine at higher speeds. Opening the battery contactor does prevent this braking torque, but results in the power switches being destroyed by excessive voltages.

A (usually) low-resistance failure of a power-switch transistor gives rise to an unacceptable braking torque at the electrical machine here as well.

A safe drive state can only then be achieved by opening the battery contactors and all-pole short-circuiting of the electrical-machine terminals. The three high-side switches or the three low-side switches are short-circuited, depending on the type of fault involved. In the event of a switching-transistor fault with a conducting failure, the neighboring switches are conductively connected; in the event of a blocking failure, the three opposite switches. It is essential to avoid a simultaneous switching-through of opposite transistors, as otherwise a short-circuit of the battery would occur if the battery contactors were not yet opened.

If a PSM machine has to be permanently short-circuited at higher rotational speeds and the vehicle is nevertheless to be moved (towed) at a not inconsiderable speed, this constitutes a considerable thermal load on the inverter – there is no short-circuit current flow.

Immunity against interference

The high rates of voltage rise at the machine windings when the terminal voltages are switched in the commutation process create in the machine capacitive crosstalk to the stator core and the rotor. These high-frequency voltages emit EMC interference (electromagnetic compatibility) to the environment and must be damped in the best way possible. Damping measures entail shielding the supply lines to the electrical machine and to the battery and filtering the direct voltage (DC) at the inverter input. DC input filtering in the simplest case consists of film capacitors to the inverter housing or to vehicle ground (Y-capacitors) to short-circuit these parasitic voltages in as short a route as possible via the ground path. Depending on the specification, the outlay can extend up to multi-stage filters of capacitors and inductors. With the pi filter, capacitors are switched between HV+ and HV–; an inductor is longitudinally switched between these X-capacitors. This produces a diagram in the form of the Greek π.

When shielding the machine supply lines it must be noted that, as a result of the different time characteristics of the phase currents, when the individual strands are shielded circulating currents can develop in the shields and across the connection elements to the inverter and the machine of a similar magnitude to the machine current. If possible, the machine supply lines should be subjected to common shielding (cumulative shield).

The voltage overshoots at the power switches during the commutation process must be taken into account in the reverse-voltage capability of the semiconductors. The higher these voltage peaks are, the higher – for a given (usually standardized) semiconductor reverse voltage – the supplying battery voltage and thus the operable drive power can be selected. For this reason a low-inductance design of the commutation circuit and if necessary voltage-limiting protective circuits (active clamping) of the power switches (transistors) provide for the best possible inverter performance.

Mounting

While for small three-phase AC drives in the vehicle (e.g., radiator fans, transmission oil pumps, electric air-conditioning compressors) the inverters are currently usually mounted directly on the end face of the electrical machine, this is not yet standard practice for the vehicle drives. The inverter is for the most part assembled as a single device or in combination with a DC-DC converter. The latter serves to provide the isolated supply of the 12 V vehicle electrical system and its consumers from the HV vehicle electrical system.

In hybrid-vehicle applications with one or two electrical machines for the drive in the main drivetrain and if necessary an additional electric final-drive unit (e.g., for hybrid vehicles with all-wheel drive) several inverters are installed in a double or triple inverter unit. In the near future, however, there is an increasing trend towards inverters mounted directly on the electrical machine or the transmission in order to spare the complex (shielded) three-phase AC supply lines to the electrical machine and its HV connections – both in electric final-drive units and in transmissions for hybrid drives.

Activation and control

Function

The driver adjusts the torque required for propulsion via the accelerator pedal. The aim of drive control is to make available according to the request of the higher-level vehicle system an electrical-machine torque that for a constant setpoint value exhibits as smooth a torque/time characteristic as possible so as not to excite the mechanical structures of the drivetrain and the vehicle to noise generation or oscillations.

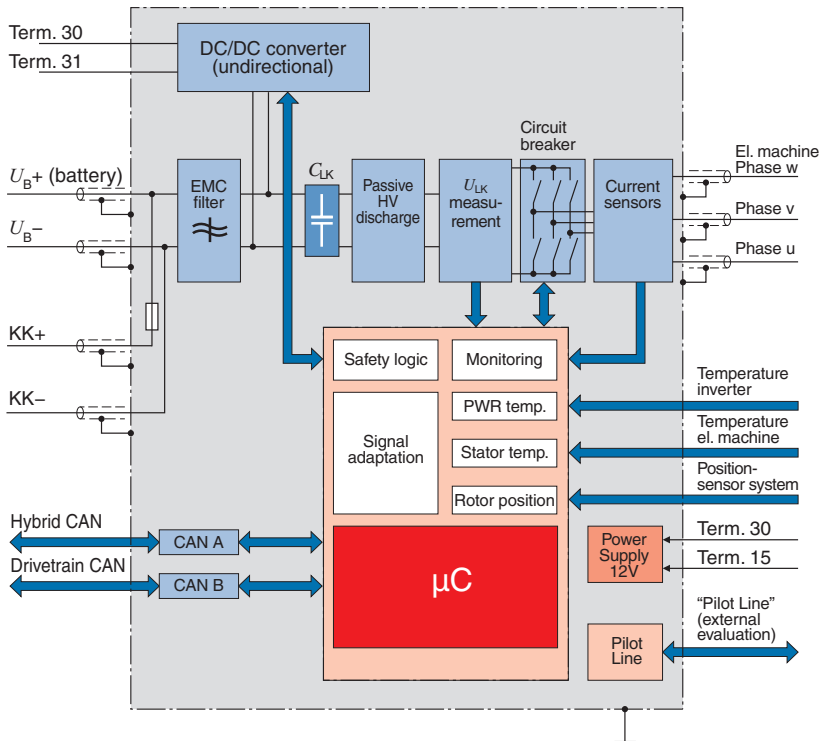
Inverter control

Inverter control is based on a central microcontroller (µC) that for control and monitoring purposes evaluates various signals (e.g., rotor position of the machine, supply voltage, machine currents and temperatures) and generates the switching signals for the power switches in accordance with the setpoint-value requests (Figure 4).

The EMC filter damps the transmission of interference voltages caused by inverter switching to the HV vehicle electrical system in order to avoid radio-reception interference. Shielding of the HV cables is a further supporting filter measure here.

Because the link capacitor stores a comparatively high amount of energy, it must in the event of a fault be discharged within 5 seconds. This is performed by a special discharging circuit or discharging

Figure 4: Circuit structure of an inverter, here with additionally integrated DC-DC converter
U_B Voltage of traction battery, C_{ZK} Link capacitor, KK Air-conditioning compressor.



function. Isolated measurements of the DC link voltage and the phase currents to the machine and of the current position of the machine rotor are required for drive-torque control. For safety reasons all three phase currents are recorded although on the condition that the total current is zero only two of these would be required.

The derating functions, i.e., the thermally based measures for power reduction, are if necessary controlled from the recording of the component temperatures or the downstream observer models.

The inverter is in communication via a number of communication interfaces (CAN or FlexRay, high-speed CAN with up to 500 kBit/s) with the vehicle's other intelligent control components, in particular the vehicle control (see VCU, Vehicle Control Unit).

The "pilot line" is a hardware wire which forms a closed conductor loop via additional contacts at each critical HV connection point or on covers which could be opened in the vehicle. Its status is monitored for integrity by way of resistance measurement by the safety technology in the battery. The battery main contactors open in the event of a fault.

The internal power supply for the control electronics is delivered from the 12 V vehicle electrical system and in the interest of safety also with an additional backup supply via a DC-DC converter from the HV vehicle electrical system. The trigger circuits (gate-driver circuits) for the power transistors, but also their monitoring circuits, must have an isolated design.

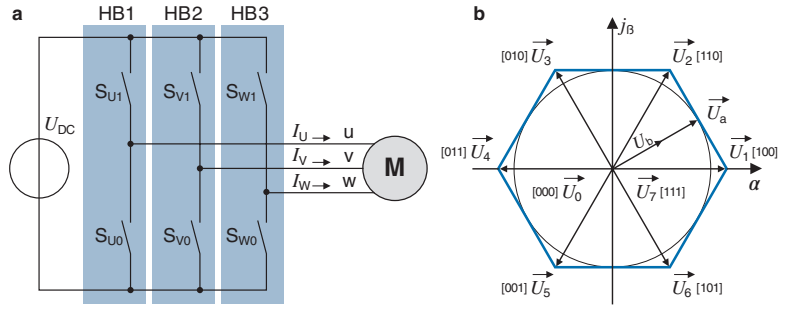
Three-phase voltage system

From the eight possible switching states of the six power switches the three phase voltages are applied at the machine terminals in the form of pulse patterns. Table 1 shows the various possibilities. The switching state 1 means that the upper switch of a half-bridge is closed. In the switching state 0 the lower switch is closed. Only one switch of a half-bridge is ever closed; two closed switches of a half-bridge would cause a battery short circuit.

The "space vector" (representative sinor) is used in the representation of a three-phase voltage system in a mathematically complex formulation . It describes in a simple mathematical notation – and illustrated on the complex level as a vector – the amplitude and the phase angle of the instantaneous value of the voltage system (Figure 5). The eight switch-position variants are thus assigned six active space-vector states and two "zero-voltage vectors" (U_0 and U_7). The states with all three half-bridges simultaneously fed through to the high-side or low-side are equivalent with regard to the machine terminals; they constitute the zero-voltage vector or terminal short circuit.

The length of the space vector describes the voltage amplitude; it is ad-

Figure 5: Switch position and space-vector assignment
a) Circuitry, b) Space vector.
U_{DC} Battery direct voltage, I Phase current, U_n Phase voltage at machine terminals, HB Half-bridge, S Switching state, M Electrical machine.



justed by means of the pulse durations of the activation. The orientation of the space vector describes the phase angle of the system; this is adjusted by pulse-duration-modulated oscillation between two adjacent switching states.

The ideal representation of a three-phase voltage system would be achieved if the vector moved continuously – without pulse-related jumps – on a circular path. This is not possible for the inverter due to the few possible discrete switching states. But because the current due to the machine inductance follows a strongly smoothed voltage, it approaches a sinusoidal shape with the frequency of the voltage fundamental component. This oscillation constitutes the machine supply frequency. The sinusoidal shape could be represented more ideally by increasing the inverter pulse frequency, but this would be accompanied by an increase in inverter losses. Today's IGBT inverters for vehicles operate with a pulse frequency of around 10 kHz. In particular drive configurations (above all in hybrid vehicles) it is also possible, for reducing losses in certain operating ranges, to work with a reduced pulse frequency, particularly when the battery voltage is adapted by upstream power-electronics controllers to the electric drive's power demand.

Control strategy

The switch-control strategy for sinusoidal supply is called SVPWM (space vector pulse width modulation). It is the basis of most control strategies. In order on the other hand to achieve further current and torque increases, i.e., better specific inverter power, it is possible however to deviate from this control strategy (overmodulation, block operation).

Control-engineering implementation

Electric vehicle drives have complex closed-loop control circuits to reduce torque and avoid drivetrain vibrations. Current control forms the innermost, fastest circuit. Coordinating this is torque control, which derives an estimated actual torque value from measurements of stator current and current rotor position and estimation-model calculations for the magnetic fluxes in the machine and the temperatures of the critical assemblies. Drive control and the real-time observer simulations required considerable microcontroller power and signal-acquisition precision. This actual torque value must be formed as precisely as possible because it not only is used for driving-comfort purposes but also forms the basis of the drive safety concept.

Working with asynchronous machines is in terms of control engineering much simpler than using synchronous machines. Synchronous machines create their torque by “displacing” the magnetic fields of a current-fed stator and a rotor equipped with magnets. It is therefore necessary here to know precisely the absolute rotor rotation-angle position. Today these are usually sensed by reluctance resolvers or eddy-current position sensors in combination with specialized evaluation-electronics modules.

Derating concept

The temperatures of critical components (e.g., the PSM rotor equipped with permanent magnets) are estimated in real

Table 1: Switching state and phase voltages

Basic voltage space vector	Switching state			Phase voltages		
	S_{U1}/S_{U0}	S_{V1}/S_{V0}	S_{W1}/S_{W0}	U_{UV}	U_{VW}	U_{WU}
U_0	0	0	0	0 V	0 V	0 V
U_7	1	1	1	0 V	0 V	0 V
U_1	1	0	0	$+U_{DC}$	0 V	$-U_{DC}$
U_2	1	1	0	0 V	$+U_{DC}$	$-U_{DC}$
U_3	0	1	0	$-U_{DC}$	$+U_{DC}$	0 V
U_4	0	1	1	$-U_{DC}$	0 V	$+U_{DC}$
U_5	0	0	1	0 V	$-U_{DC}$	$+U_{DC}$
U_6	1	0	1	$+U_{DC}$	$-U_{DC}$	0 V

time by means of observer models using measured values and other available data in order to operate the drive components in their safe operating ranges. If critical limits are reached, limitation strategies intervene to reduce loads gradually and thereby prevent damage to the components.

Closed-loop control-circuit influences
Closed-loop control-circuit influences for taking account of the elasticity of specific drivetrains (subject area of comfort) are already in use today. It is foreseeable in future that, in the case of high-power electrifications (typically electric vehicles) with the electric drive heavily involved in braking, the braking functions will also in part be integrated in drive control in order to avoid communication dead times.

Software and functional safety

Requirements
The inverter software must satisfy requirements that are comparable with those of an internal-combustion engine. As well as a vehicle immobilizer and safeguarding against unintended movement and acceleration by the three-level concept (comparison with EGAS concept for gasoline-engine control), increased functional-safety requirements (ISO 21212, [1]) with regard to the safety objective "unintended torque" however must also be satisfied. The ASIL levels B through D (Automotive Safety Integrity Level) must be realized here, depending on vehicle size, drive power, front- or rear-wheel drive, and if necessary the use of sufficiently safe signals from other ECUs in the vehicle. These safety levels are essentially achieved by redundancies in the sensor system, the supply, and the software, if necessary also with diversity. In other words, not only is the system redundant in design, but also different solutions are used.

ASIL is a risk-classification scheme which is defined in the standard ISO 26262 [2].

Three-level concept
The three-level concept (Figure 6) monitors on a lower software level (level 2) the plausible function of the actual control software and the functional software (level 1), whereby it continuously monitors the ranges of values in which the variables move. At a further lower monitoring level (level 3) the microcontroller (μC), its memories, and the function of the sensors are monitored, whereby test operations are fed in. An error counter in the assigned watchdog hardware module aggregates registered errors and if necessary triggers the "safe state" of the electric drive. This is currently defined in the case of a PSM machine as its three-phase all-pole short circuit.

Figure 6: Principle of the three-level concept for ECU monitoring [3]

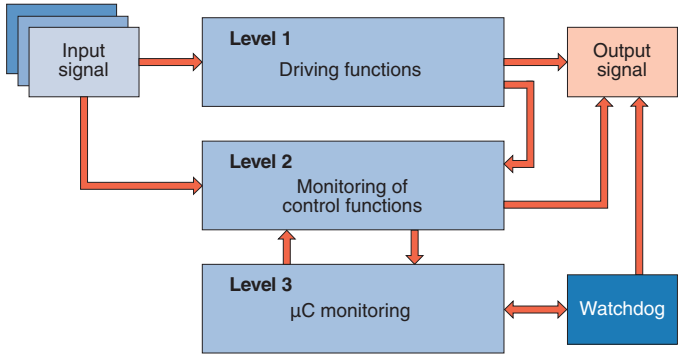
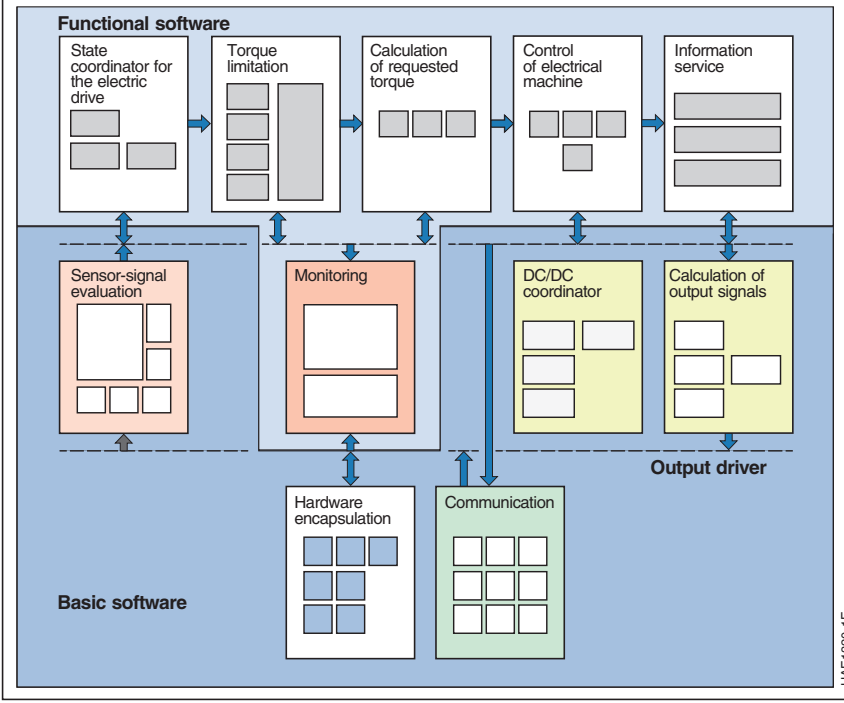


Figure 7: Software-structure design



Functional safety
Not only the ECU computer is monitored within the framework of functional safety. Also – depending on the requirement level – hardware redundancies and diversities are introduced in the components used. This does not necessarily have to be realized in an ECU (here in the inverter). It is perfectly possible to realize these functionalities on the vehicle level by combining the various ECUs and their data. The aim of monitoring is to state definitively that the actual torque of the drive is in a permitted tolerance band for the current operating point around the setpoint value.

In addition to the functional software for the drive which is subject to monitoring here, further program sections are required for time-critical reading-out of sensors and activation of the power switches (the so-called complex drivers), for communication and the hardware-related basic software for operating the ECU (Figure 7). The functional software can come from a single source or in parts from different sources (software sharing by vehicle manufacturer and different suppliers). However, this calls for the appropriate level of co-ordination and each supplying party must provide its specific monitoring function. The basic software is supplied in each case by the equipment manufacturer.

Equipment mechanics

Mounting the inverter
Today's inverters are for the most part tested devices with cable connection or designed for directly mounting on the electrical machine or transmission housing. A market for tested inverter assemblies seems to be emerging. In which partitioning and installation situation an inverter is used is in part specific to each vehicle manufacturer. Standardization in the market is difficult to conceive.

Cooling
The loss-generating inverter assembly (essentially the power modules and the link capacitor) is in most cases liquid-cooled with coolant mixed with glycol. The typical volumetric flow is usually max. 10 liters per minute, the maximum temperature depending on the application is between 65 °C and 85 °C.

Inverter design concepts
The heart of the inverter, the power-switch module, comes in a variety of designs.

Bond-frame module
The bond-frame module is sintered with semiconductor chips or soldered to an insulating ceramic metallized on both sides (AMB, Active Metal Brazing; DCB, Direct Copper-Bonding; made of Al_2O_3, AlN or Si_3N_4). The top surfaces of the chips are contacted with thick-wire bonds or with soldered or sintered flat conductors. The entire module is passivated with gel.

Mold module
The semiconductor chips are soldered or sintered to lead frames, the top surfaces are contacted with lead frames. The entire module is packed with transfer mold. Owing to the size, it is for the most part designed as a half-bridge module, but sometimes also as a single switch. Insulation from the cooler is provided by laminate or intermediate ceramic layers.

Cast module
This module has the same design as the mold or the bond-frame module, but is passivated by synthetic-resin casting compound.

Soldered bond-frame modules
This technology is no longer state of the art in that the risk of solder fatigue significantly limits service life and load capacity.

Low-temperature silver sintering
Alternatively, there is a trend in chip bonding towards low-temperature silver sintering with significantly longer service life. Mold and casting technologies likewise offer significantly longer service life for soldering joints.

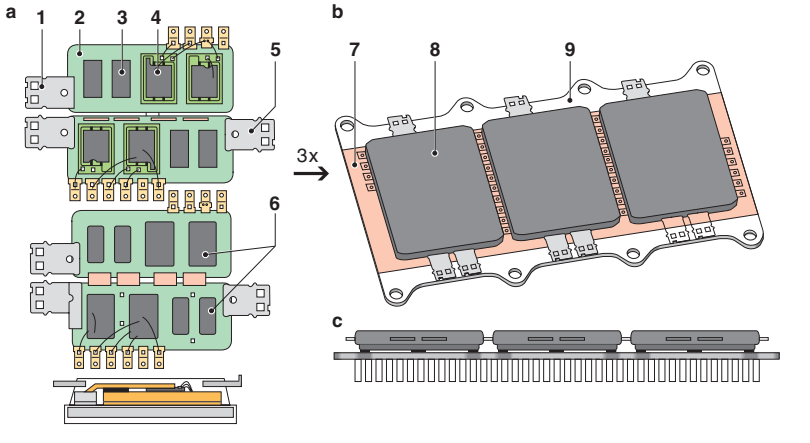
Heat dissipation
Pin-fin concept
Virtually all of today's power modules make use of a pin-fin heat sink made of aluminum or copper for heat dissipation. These are baseplates onto which the power modules are mounted and whose other side projects with so-called pin-fins (finger-shaped surface-area enlargements and turbulence generators) into the water cooling circuit (Figure 8). Related measures to improve heat dissipation are used in other cases.

Because an inverter uses power from 100 to 200 kW at 96 to 98 % efficiency, losses of several kilowatts must be dissipatable.

Cooler types
Two basic types of cooler are currently in series production:
– The water path is created in the aluminum housing and sealed with the pin-fin on the other side.
– When the cooler is closed the pin-fins are also connected with the other side of the water path. This cooler has the advantages of higher pressure resistance and better heat dissipation.

Design
Link capacitor
The link capacitor is likewise affected by losses on account of its dielectric losses and transmission of the direct current to

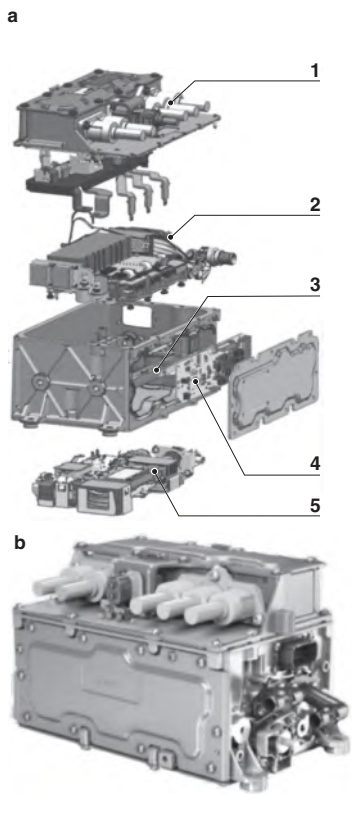
Figure 8: Design of a half-bridge power module and inverter power module on a pin-fin cooler plate
a) Baseplate and top-surface clip of a half-bridge power module,
b) Inverter power module with three half-bridges,
c) Inverter power module on pin-fin cooler plate.
1 U_B+ and U_B− (battery-voltage connections), 2 Copper baseplate, 3 Diode semiconductor chips, 4 IGBT semiconductor chips, 5 Phase connection, 6 Top-surface clip, 7 Insulating layer, 8 Half-bridge power module, 9 Pin-fin cooler plate.



the power modules (primary current flows via conductor bars in close proximity to the capacitor or if necessary also through the capacitor busbar). It is advantageous to run the primary current past the capacitor elements over as short a route as possible and essentially only to apply the ripple current to the busbar of the elements. Capacitors for inverters in the upper power range have for the most part a connection to the cooling medium (Figure 9).

Figure 9: Mechanical design of a combination of inverter and DC-DC converter

Assemblies are placed at both ends of an internal cooler.
a) Exploded view,
b) Equipment view.
1 HV connections DC and AC, HV-DC for small drive (AC compressor),
2 Inverter assembly with pin-fin water cooler, power modules, link capacitor, gate-driver circuits, phase-current measurement,
3 Liquid cooler,
4 Microcontroller control circuit,
5 DC-DC converter.



Commutation circuit
The busbar of the commutation circuit between the link capacitor and the DC inputs of the power module must be designed to be as low-inductance as possible, i.e., the potential surfaces must be routed as closely as possible to each other and the current paths during commutation between the switches should encompass as few surfaces as possible.

Gate-driver circuit
The gate-driver circuit is placed as close to the power switches as possible, but the electromagnetic fields for the most part must be shielded in the area of the power modules so that there can be no interference with the control circuits even in the event of extreme transient currents.

HV connections
The HV connections also deserve special attention since they too are affected by losses and adequate heat dissipation must be provided for or the creation of heat loss should be avoided as far as possible (screw connections are advantageous).

References
[1] ISO 21212: Intelligent transport systems – Communications access for land mobiles (CALM) – 2G Cellular systems
[2] ISO 26262: Road vehicles – Functional safety.
[3] https://www.autotec.ch/technik/pdf/ms_E-GAS_Ueberwachung.pdf

DC-DC converters

Application

A DC-DC converter, also called a voltage transformer, is an electric circuit that converts a direct voltage into a different direct voltage [1]. Conversion serves to
– adjust a different voltage level,
– adjust a stable voltage,
– facilitate a variably adjustable voltage or
– electrical isolate two voltage supply systems.

A number of these requirements can be simultaneously achieved with a voltage transformer.

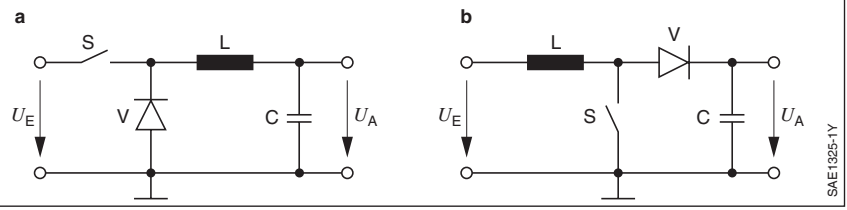
Voltage levels
Conversion of high to low voltages
Every vehicle has for power-supply purposes a 12 V electrical system, consisting of a current source, an energy accumulator/storage device (battery), and the distribution (cables). This is also the case for electric and hybrid vehicles, which have a traction battery with a voltage level of 400 V or 800 V.

Whereas an alternator serves as the current source in conventional vehicles with internal-combustion engines and solely a 12 V vehicle electrical system, a supply from the traction battery via a DC-DC converter is used in electric and hybrid vehicles. This system converts high direct voltage (HV) into low direct voltage (LV) for the 12 V vehicle electrical system. Power ranges typically between 2.4 and 3.5 kW; peak demands over and above this power output are buffered by the 12 V battery (predominantly a lead-acid battery, in future also a lithium-ion battery). The advantages of a supply from the traction battery via the DC-DC converter compared with the alternator in a conventional 12 V electrical system are higher efficiency and a supply of energy that is not dependent on engine speed.

Conversion from 48 V to 12 V
In mild-hybrid vehicles the 12 V electrical system is supplied from the 48 V battery. Therefore voltage transformation with a DC-DC converter is also required here.

Requirements
DC-DC converters are subject to different requirements on account of the voltage levels. There are therefore essentially two distinct categories: non-isolating (electrically connected) and isolating (electrically isolated) DC-DC converters.

Figure 1: Diagrams of electrically connected DC-DC converters (block diagrams)
a) Forward converter (step-down converter),
b) Blocking converter (step-up converter).
U_E Input voltage, U_A Output voltage.
S Switch, V Diode, L Coil (inductor L), C Capacitor (C).



Non-isolating DC-DC converters

Basic design and operating principle
DC-DC converters essentially consist of switches, diodes, and energy storage devices (inductors or capacitors, or both). There are various ways to wire up these elements and there are several ways of classifying them. Some arrangements have gained acceptance in practice, and these will be described in the following.

Electrically connected DC-DC converters are the simplest type, most notably the step-down converter and the step-up converter (buck converter and boost converter).

With reference to these two simple converters a further distinguishing feature may be introduced: forward converter and blocking converter. The step-down converter in Figure 1a is called a forward converter because it transfers energy to the output during the switch's conducting phase. With the blocking converter in Figure 1b, on the other hand, the energy transfer to the output takes place during the switch's blocking phase.

Step-down converter
In the case of the step-down converter (Figure 1a) the switch S can assume two states: switched on ("on") and switched off ("off"). Power-electronics components such as MOSFET and IGBT are used as switches. The switch S can conduct either the input voltage U_E or no voltage to the subsequent circuit. Taking the repetition period of a switch-on and switch-off operation as duration T (period duration, and derived from this the switching frequency) and keeping this constant, a square-wave voltage is obtained whose components U_E and 0 are dependent on the duty ratio t during a period. The ratio of t to T is defined as the duty cycle D and provides the relationship between input and output voltages.

However, the voltage downstream of the switch has a non-constant square-wave shape and must still be changed to direct voltage. This task is performed by the two energy-storing elements, namely the coil (inductor L) and the capacitor (C), which in their arrangement act as a second-order low-pass filter. The switch S interrupts and conducts the current alternately, which results in a discontinuous flow of current. Discontinuous flows of current through inductors result in high induction voltages. In order to avoid these high induction voltages, the current must flow continuously through the inductor. This is ensured by a free-wheeling diode. The is the diode V, which ensures that, with the switch S in the off state, the current can flow via the inductor L, the load (not shown in Figure 1) in the output circuit and C via V back to the inductor L. The diode V blocks when the switch S is closed. The capacitor C serves to smooth the ripple output voltage. Consequently, the low-pass filter generates from L and C the mean value from the chopped DC input voltage downstream of the switch S. The output voltage is obtained as the Dth part of the input voltage:

$$U_A = D\,U_E, \text{ mit } D = \frac{t}{T}.$$

In this way the input voltage U_E is converted to a lower voltage U_A.

Step-up converter
Here too the switch S can assume two states: switched on and switched off (Figure 1b). In this way the switch S can apply the input voltage U_E to the inductor L, which results in a build-up of current in the inductor. The diode V is blocked in the process. During this phase the capacitor C at the output maintains the voltage and supplies current to the load circuit. When S is switched off, the inductor drives the current with the energy from the previously built-up magnetic field further across the diode V to the output. The inductor is now with its induced voltage in series with the input voltage and thus the output voltage U_A is higher than the input voltage:

$$U_A = \frac{1}{1-D}\,U_E.$$

48 V-12 V voltage transformers

DC-DC converters are used in vehicles with 48 V and 12 V electrical systems to connect the two electrical subsystems to each other. These voltage transformers are therefore always bidirectional and can work in both directions

– from 48 V to 12 V as step-down converters and
– from 12 V to 48 V as step-up converters.

Because no electrical isolation is required on account of the voltage of 48 V, the simple topology of the step-down and step-up converters is preferred (see Figure 1). An active rectifier with a MOSFET is used instead of a passive rectifier with a diode. MOSFETs have an inherent source-drain diode (also known as a body diode) which can be used for rectifier purposes. If the MOSFET is actively switched on during the diode's conducting phase, the losses in the now “active” rectifier decrease. Consequently, the MOSFET is now able to perform the two “switch” and “diode” functions, depending on its activation.

Figure 2a shows the basic circuitry of such a DC-DC converter with two MOSFETs arranged as a half-bridge. This circuitry can also be seen in Figure 3. For the step-down converter the upper MOSFET is used as the switch, the lower MOSFET operates as the diode (Figure 2b). This arrangement corresponds exactly to Figure 1a.

For the step-up converter the lower MOSFET operates as the switch, while the upper MOSFET operates as the diode (Figure 2c). This corresponds to the mirrored arrangement of Figure 1b.

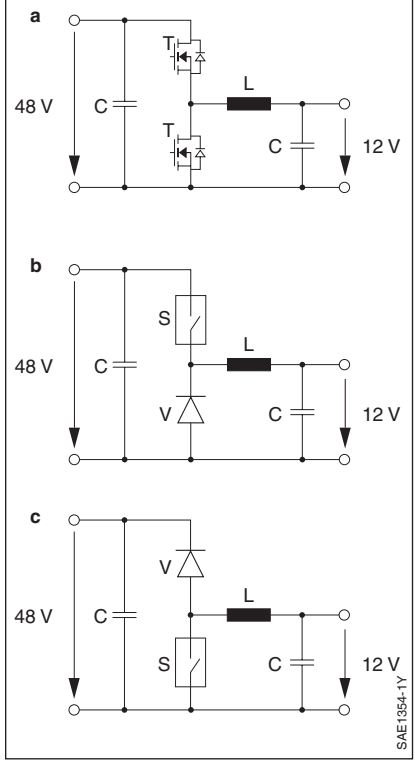
In other words, the DC-DC converter can operate from 48 V to 12 V as a step-down converter and from 12 V to 48 V as a step-up converter. Only the switch and rectifier functions change between the two MOSFETs.

For higher currents a number of phases are connected in parallel to create multiphase converters. These operate with the same switching frequency, but with a phase shift. This on the one hand splits the power and on the other hand delivers benefits through the superimposition of the voltage ripples of the phases with the result that the converter switching frequency virtually increases. Smaller voltage ripples and benefits in filtering are thereby obtained.

Figure 3 shows a possible topology. One phase consists of two MOSFETs and one inductor. The “filter” functions are jointly used by all the phases. Even the safety switches which isolate the converter from the vehicle electrical system in the event of a fault are jointly used and therefore feature only once per side. There is if nec-

Figure 2: Function of a bidirectional voltage transformer

a) Design with two MOSFETs,
b) Function as down controller from 48 V to 12 V,
b) Function as up controller from 12 V to 48 V.
T MOSFET, C Capacitor, L Coil,
V MOSFET functioning as a diode,
S MOSFET functioning as a switch.



essary also polarity reversal protection on 12 V side of the converter. The control (or closed-loop control to be exact) ensures that the setpoint values for the respective output voltage are constantly set even under varying load conditions and fluctuating input voltage. The duty cycle of the half-bridges is set accordingly for this purpose.

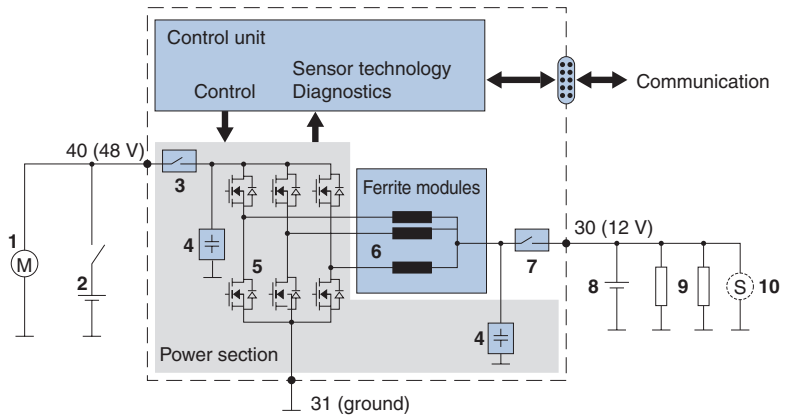
Traction-electrical-system DC-DC converters

Traction-electrical-system DC-DC converters are likewise non-isolating converters and usually operate with the same topology as 48 V-12 V voltage transformers. They are used to adapt the voltage in the traction electrical system of a vehicle; this is the HV vehicle electrical system (high voltage) between the traction battery and the inverter. The traction battery has a voltage that is dependent on the state of charge; this has a noticeable impact on the maximum available drive power. To prevent this, the traction-electrical-system DC-DC converter regulates the inverter input voltage to a stable voltage, regardless of the changing battery voltage.

Because traction-electrical-system DC-DC converters are designed for a much higher power range than 48 V-12 V voltage transformers (> 50 kW), these are set up with IGBTs (with parallel-connected freewheeling diodes) and operated with lower switching frequencies than 48 V-12 V converters.

Figure 3: 48 V-12 V voltage transformer

1 Electrical machine (48 V), 2 48 V battery,
3 Safety switch (48 V side),
4 Filter function and voltage smoothing with capacitor,
5 Three half-bridges each with two MOSFETs, 6 Coils (inductors),
7 Safety switch (12 V side), 8 12 V battery, 9 12 V consumers,
10 Starter (optional, in a control situation the electrical machine performs the starting operation).
30, 31, 40 Terminal designations.



Isolating DC-DC converters

DC-DC converters of high voltages to 12 V

Design

For safety reasons the traction electrical system in the vehicle is designed as an electrically isolated network (floating network) with no connection to vehicle ground. Energy transfer from the traction electrical system to the 12 V vehicle electrical system is therefore always performed with electrical isolation. This prevents dangerous situations in the event of single faults in the system. A transformer is used in the voltage transformer for electrical isolation. Full-bridge converters are usually used for this performance class. Figure 4 shows the basic design of an electrically isolated DC-DC converter.

Operating principle

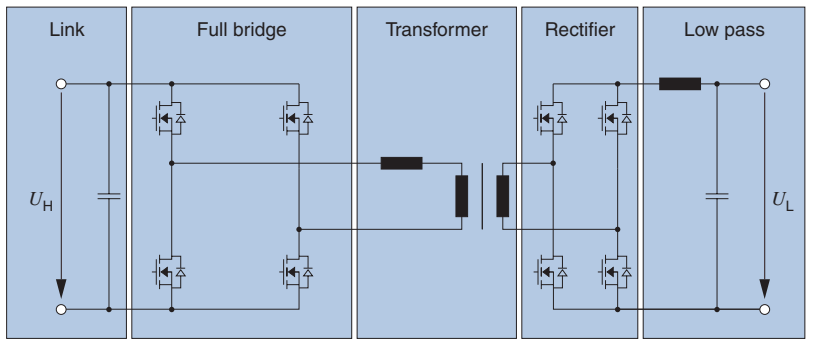
First the input direct voltage (high voltage) is “chopped”, i.e., converted into an alternating voltage, by the full bridge. The transformer transmits the alternating voltage from the primary side to the secondary side, in the course of which the voltage can be adapted via the turns ratio between the primary and secondary coils. A step-down ratio of 1:10 to 1:13 is customary here. Now the adapted, electrically isolated output voltage is applied as alternating voltage at the transformer's secondary coil. This alternating voltage is converted by the rectifier into a pulsating direct voltage and smoothed by the output filter (LC filter) to become the output direct voltage.

The output voltage is set to the desired value by a modulation of the full bridge. There are two modes to choose from for the modulation:

– PWM activation, whereby the output voltage is adjusted via the duty cycle and
– phase shifting, whereby each half-bridge is activated with a duty cycle of 50 % and the output voltage is adjusted via the phase relation of the two half-bridges.

Both methods have benefits and drawbacks; phase shifting has gained acceptance for isolating DC-DC converters in vehicles. Accordingly, this method is referred to as the phase-shifted full-bridge (PSFB). A key advantage of the PSFB is that the leakage inductance of the transformer (shown in Figure 3 in series with the transformer's primary coil) can be used together with the 50 % operation of the half-bridges for a soft-switching operation of the DC-DC converter. This lowers the switching losses of the semiconductors so that efficiency and switching frequency can be increased. A higher switching frequency has a favorable impact on the installation space and the costs of the passive components.

Figure 4: Design of an electrically isolated DC-DC converter
U_H High voltage,
U_L Low voltage.



Typical switching frequencies range between 70 and 100 kHz and sometimes also around 140 to 145 kHz. The choice of these frequencies avoids the range around 125 kHz in which radio systems for vehicle access or tire-pressure monitoring frequently operate. From 150 kHz upwards the restrictions imposed by EMC regulations governing emitted interference (electromagnetic compatibility) take effect such that the switching frequency should be much lower.

Further functions of the DC-DC converter in the vehicle

Most of the methods used are also capable of reverse operation and consequently can supply energy from the 12 V vehicle electrical system to the traction electrical system. This is used for two functions. The first function is for precharging the capacitors in the traction electrical system; this saves on costs and installation space for the charging contactors in the traction battery. The second function of reverse operation is the test pulse for the 12 V battery. Here the 12 V battery is briefly subjected to a discharge current; the HV battery absorbs the energy as charging current. From this the battery sensor determines the current performance of the battery.

Future functions will promote the redundancy of the energy supply to the 12 V vehicle electrical system so that safety-relevant consumers and their function – in spite of a single fault in the DC-DC converter – can continue to be supplied with energy at least for a certain amount of time.

DC-DC converters of high voltages to 48 V

DC-DC converters with input voltages of several hundred volts and an output voltage of 48 V are very similar in their design to DC-DC converters with a 12 V output. Essentially the transformation ratio of the transformer is changed. Furthermore, the secondary-side circuits must be adapted to the higher output voltage. The rectifiers need a higher reverse voltage, the filter capacitors need higher dielectric strength, and the sensors for current and voltage must be adapted to the other current and voltage.

Because primarily current is a crucial design parameter for DC-DC converters, higher output power can be achieved with little additional outlay thanks to the higher voltage compared with the 12 V DC-DC converter.

References

[1] U. Schlienz: Schaltnetzteile und ihre Peripherie. 6th Ed., Verlag Springer Vieweg, 2016.

On-board charger

Application

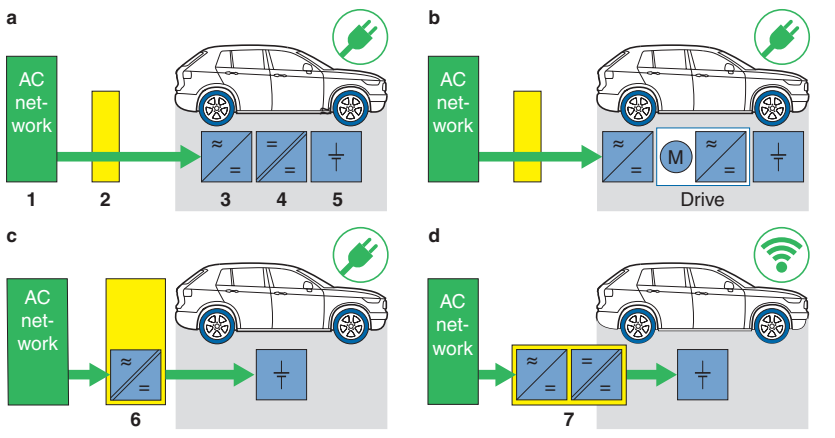
The component in externally rechargeable hybrid, fuel-cell and electric vehicles which is responsible for charging the traction battery is called the charger. All externally rechargeable vehicles are typically equipped with an on-board charger which is carried – permanently installed – in the vehicle. The connection to the power-supply network (infrastructure) is established via a charging station with the aid of a charging cable, which is why such charging concepts are also called conductive chargers (Figure 1).

Alternating-current charging

The public low-voltage network serves as the infrastructure for today's conductive charging concepts because it is comprehensively available and thus can be used directly. The basic design of an on-board charger with AC interface (alternating current) between the vehicle and the power-supply network is illustrated in Figure 1a. Alternating current is supplied to the vehicle (hence also the term AC charging); the voltage is transformed from alternating current to direct current in the on-board charger inside the vehicle. The traction battery can be charged with this direct current.

Figure 1: Basic design of on-board chargers

a) AC charging (standalone charging concept, additional charger in the vehicle),
b) Integrated charging concept, drive components used to charge the traction battery,
c) DC charging (external charger, the traction battery is directly charged with direct current),
d) Inductive charging (external and internal charger in the vehicle with inductive coupling).
1 Public AC network, 2 Charging station (e.g., charging post, wallbox),
3 Voltage transformer (from alternating voltage to direct voltage, non-isolating without electrical isolation),
4 DC-DC converter (from direct voltage to direct voltage, electrically isolated with electrical isolation),
5 Traction battery, 6 Charging post for DC charging with integrated charging electronics,
7 Inductive charger (transformer primary coil of the DC-DC converter in the charger, secondary coil in the vehicle).
M Electrical machine, DC Direct current, AC Alternating current.



Integrated charging concept
Aside from standalone chargers there are also charging concepts in which components already installed in the vehicle (e.g., inverter or electrical machine) are used for the charging process (Figure 1b). Because for the most part the inverter is the integral component of such concepts, the term inverter charging is also used in this context. Such chargers are however generally not electrically isolated and for that reason they will not be discussed further here.

Direct-current charging
Direct-current (DC) charging plays a central role in the area of fast charging. Here the actual charging electronics is installed in the charging station (Figure 1c), and for that reason weight and vibration-strength requirements are less relevant. The charging station ascertains the desired charging current by communicating with the vehicle. Wires carried in the charging cable are used for communication.

In future DC networks (micro-networks), such as those featured for example in companies, could be used for charging. This could dispense with rectification inside the charger. The charging stations contain a DC-DC converter which adapts the charging voltage to the vehicle demand.

Inductive charging
For the sake of completeness it should be pointed out that engineers are already working on inductive chargers for electrified vehicles in which the energy transfer to the vehicle takes place via a split transformer between infrastructure and vehicle (Figure 1d) and the charging process can start automatically with the vehicle appropriately positioned on the charger. This transformer assumes the function of the transformer in the DC-DC converter (see DC-DC converters). The primary coil is located in the charger, the secondary coil is situated with the rectifier in the vehicle.

Requirements

The fundamental requirements imposed on an on-board charger are laid down by the standards to be observed on the network side and by the vehicle-side battery and safety-concept requirements.

Network-side requirements
To ensure that a charger can be used on a worldwide scale, a voltage range of 85 to 265 V (with 50 or 60 Hz) is typically provided for. It is important to take into account that typically a power adjustment is made under the nominal voltage of 230 V.

The current harmonics to be maintained in the public network (IEC 61000-3-2 [1]) constitute a further network-side requirement which has a direct influence on the charger topology. These necessitate the use of an active power-factor correction (PFC, see Power-factor correction).

Vehicle-side requirements
On the vehicle side above all the traction-battery requirements must be taken into consideration. Aside from the voltage range to be covered, stretching in current electric vehicles from 240 V to 450 V (for passenger cars), the charger should provide as smooth a charging current as possible, i.e., with low ripple. This requirement plays a central role in single-phase chargers because the charging current exhibits a strong ripple if the pulsating power consumption from the AC network is not smoothed.

Aside from the battery requirements, the safety concept of a charger plays a central role in its integration in the vehicle. Electrically isolated on-board chargers are state of the art. The power-supply network and the vehicle can be decoupled by means of electrical isolation, which makes it easier to bring potentially occurring insulation faults under control. This is because the vehicle body is connected to the ground conductor of the power-supply network during the charging process. It should basically be pointed out that electrical isolation proves to be disadvantageous with regard to achievable efficiency.

Future requirements
Aside from the classic traction-battery charging function or providing electrical energy for example for vehicle climate control prior to starting a trip, further functions will gain in importance in future:
– V2L (Vehicle to Load),
– V2H (Vehicle to Home),
– V2G (Vehicle to Grid).

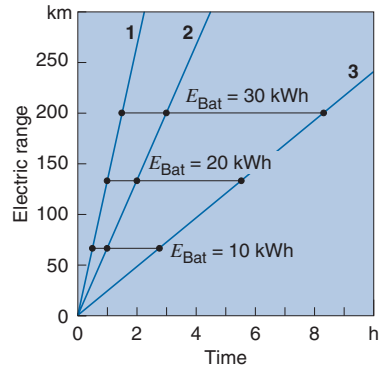
What these new functions have in common is the required bidirectional energy transfer of the charger. This means that future chargers must be able to transfer energy not only from the supply network to the traction battery but also from the traction battery back to the network. While V2L merely permits the operation of a simple consumer (e.g., electric grill) from the vehicle, V2H is concerned with the intelligent integration of the vehicle storage device in the existing domestic electrical installation in order to be able, by including the vehicle battery, to better utilize for example the electrical energy self-generated by a photovoltaic system.

The concept with the greatest complexity is V2G, with which parked vehicles make available some of their battery capacity for appropriate payment in order to participate in the energy control market so as ultimately to get rid of the expensive peak-load power stations and at the same improve the stability of the power-supply system through a better balance between generation and consumption. For the V2G approach to work, sufficient numbers of vehicles must always be connected to the network, which presupposes an appropriate degree of market penetration by electromobility.

Charging power

The available charging power is very important since its has a direct bearing on the rechargeable energy in a specified time. The higher the charging power, the greater the recharged energy to the traction battery during the charging time and hence the possible range with the electric drive. While it takes just a few minutes to refuel a conventional vehicle with an internal-combustion engine, it can take several hours to charge electrified vehicles, depending on the available charging power and the battery capacity. To illustrate this, the electric range that can be achieved with the recharged energy (i.e., the range under electric-only driving conditions) is plotted against the charging time for different charging-power figures in Figure 2. Around 1.5 hours are needed even with a charging power of 10 kW in order to be able to recharge for example the energy for a range of 100 km. The battery and charger losses are disregarded in the idealized treatment given here. It

Figure 2: Electric range as a function of charging time and charging power
Graph applies to a vehicle with 1 t weight and a specific energy demand of 150 Wh/km.
1 P_B = 20 kW,
2 P_B = 10 kW,
3 P_B = 3.6 kW.
P_B Charging power,
E_B Energy content of battery.



is furthermore assumed that the vehicle can be continuously charged with the full charging power.

The charging-power figures of today's on-board chargers are based on the country-specific conditions of connection to the public low-voltage network. They utilize from one up to all three network phases of the existing three-phase AC networks in the power-supply network. The charge power here ranges from 3.6 kW (single-phase, 230 V, 16 A) through 7.2 kW (single- or two-phase) up to 10 kW (single- or three-phase, 400 V, 16 A) or 20 kW (three-phase, 400 V, 32 A).

Basic design and operating principle

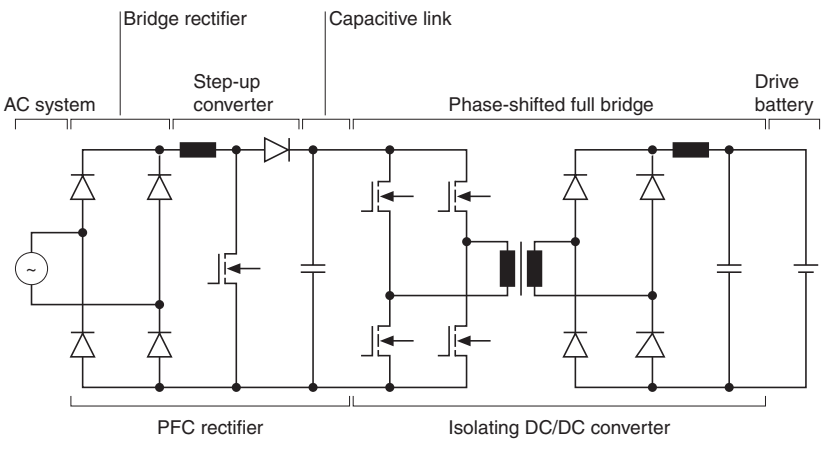
The basic design of a charger is explained in the following with reference to a single-phase charger. Figure 3 shows a state-of-the-art two-stage charger consisting of a step-up-converter-based PFC rectifier (see Power-factor correction) with a downstream isolating DC-DC converter.

The fundamental operating principles of a step-up converter and of an isolating DC-DC converter are not described here (see DC-DC converters). Rather, the specifics of the established circuit topologies when used in a charger are elaborated here.

Power-factor correction

Conventional rectifiers with a downstream smoothing capacitor are, in view of their non-sinusoidal current consumption, referred to as nonlinear consumers. The line currents of such rectifiers are composed of the superposition of higher-frequency current harmonics, which results for example in higher losses in the line transformers, but also negatively influences the compensation of the circulating currents in the neutral conductor of three-phase AC networks. To avoid these problems, it is necessary to utilize an active pow-

Figure 3: Charger in the vehicle
PFC Power-factor correction.



er-factor correction which adheres to the legally stipulated limit values for current harmonics in the range of 100 Hz to 2 kHz and makes possible a power factor (ratio of effective to apparent power) close to the value 1.

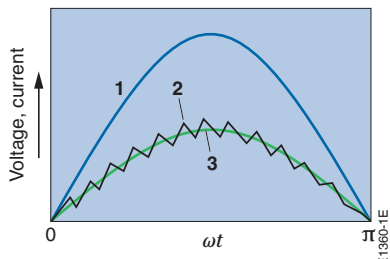
The PFC rectifier presented here consists of a conventional step-up converter with an upstream bridge rectifier which rectifies the network-side alternating voltage. For the most part step-up-based circuit concepts are used as power-factor-correction filters since these exhibit advantages over step-down systems due to their continuous input current (low filter outlay). Because of the step-up-converter approach, the DC link voltage must always be greater than the rectified line voltage. In conjunction with the 230 V AC network a DC link voltage of around 400 V is obtained which is referred to in the following as the input voltage for the subsequent DC-DC converter stage.

Because of the constantly changing input voltage the duty cycle D of the step-up converter must be continuously tracked:

$$D = 1 - \frac{\hat{U}_{pk}}{U_{ZK}} \; |\sin(\omega t)|,$$

$\hat{U}_{pk}$ = line-voltage amplitude,
U_{ZK} = DC link voltage,
ω = angular frequency of line frequency
= 2π · line frequency.

Figure 4: Network-side current and voltage characteristics of the PFC rectifier
1 Rectified line voltage,
2 Please current through step-up choke,
3 Current from the network, equal to the mean current through the step-up choke.



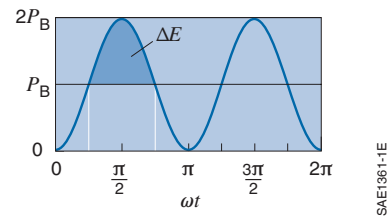
The network-side current characteristic of the PFC rectifier is illustrated for a continuous choke current (CCM, continuous conduction mode) in Figure 4. The triangular shape of the switched choke current is also shown in addition to the direct proportionality between the rectified line voltage and the mean current through the step-up choke. Typical switching frequencies for a PFC rectifier are in the range of 100 kHz in order on the one hand to be able to have compact passive components and on the other hand to achieve good efficiency in conjunction with hard-switching operation (switching losses).

The interference generated by the high-frequency switching of the PFC rectifier is suppressed with the aid of a passive network filter, which for reasons of clarity is not shown in Figure 3 [2].

Power smoothing

Located at the output of the step-up converter is the DC link, which is designed as a capacitive energy storage device with the aid of electrolytic capacitors. The need for power smoothing in single-phase chargers can be derived directly from the requirements of a line-frequency-free charging current. Because constant power cannot be drawn from the AC network, without power smoothing the pulsating power would be transferred to the output. Generally, it must be borne in mind that a constant power consumption is first possible in conjunction with a three-phase AC system.

Figure 5: Idealized time characteristic of the pulsating power consumption from the AC network
P_B Charging power,
ΔE Energy when charging and discharging the DC link.



The characteristic of the pulsating system power at the AC network is shown by way of example of a sinusoidal line current in phase with the line voltage in Figure 5. It must be recognized that the power input pulsates between zero and twice the mean charging power P_B with double the line frequency.

The pulsating system power is smoothed in the DC link by the input and output of the excessively or insufficiently delivered power, whereby the link capacitor is continuously charged and discharged. The storage and discharge of the energy ΔE in the link capacitor leads to a voltage ripple in the DC link which must be taken into account in the design of the subsequent DC-DC converter.

Electrical isolation

An overview of isolating DC-DC converters and their operating principles is provided elsewhere (see DC-DC converters). Reference is made at this point, as shown in Figure 3, to the familiar topology of the phase-shifted full-bridge (see Phase-shifted full-bridge). But also other soft-switching switching topologies, such as for example the LLC resonant converter [2], are suitable for use as isolating converters in a charger.

The function of the isolating DC-DC converter at the charger output is, aside from decoupling the vehicle from the supply network, to provide as smooth a charging current as possible to the traction battery. To this end the converter sets the requested charging current from the DC link voltage stipulated by the PFC rectifier.

Three-phase chargers

On advantage of the analyzed single-phase charger is the possibility of three-phase enhancement. Owing to the electrical isolation utilized a number of such designed chargers can be connected in parallel at the output and operated from different network phases so that a three-phase system is obtained by wiring up a single-phase system accordingly.

But, essentially, pure three-phase circuit systems also exist. In the case of such systems it must be noted that the circuit topology, aside from providing three-phase operation, must typically also be able to charge from the single-phase AC network because currently access to the three-phase AC network across the board cannot be assumed.

References

[1] DIN EN IEC 61000-3-2 VDE 0838-2:2019-12: Electromagnetic compatibility (EMC), Part 3-2: Limits – Limits for harmonic current emissions (equipment input current <= 16 A per phase).

[2] U. Schlienz: Schaltnetzteile und ihre Peripherie: Dimensionierung, Einsatz, EMV. 7th Ed., Verlag Springer Vieweg, 2020.

Batteries for electric and hybrid drives

Requirements

Hybrid and electric vehicles require a battery as an energy accumulator or storage device to operate the electrical machine. The traction battery in modern electric vehicles often accounts for more than two-thirds of the weight, size and cost of all components in the electric drivetrain. Therefore, the main challenge for the future is to improve and optimize the battery in terms of these three criteria.

There are essentially three different voltage ranges. In the LEV segment (Light Electric Vehicle) the nominal battery voltage is up to 48 V, providing an achievable power output of around 30 kW. 48 V batteries are also used for mild-hybrid systems in passenger cars. Batteries for hybrid and electric vehicles in the passenger-car sector are designed for voltages up to 400 V and can provide a power output of up to 450 kW. In the commercial-vehicle sector heavy vehicles require even higher power outputs. The DC link voltage is therefore increased to 800 V, making possible a power output of around 800 kW. Because increasing the battery voltage makes reduced charging times possible, some passenger-car manufacturers also opt for 800 V technology.

A traction battery is built from different components, not all of which need to be located in the same housing. The so-called battery pack consists of many cells, a sensor system, and a monitoring unit (BCU, battery control unit). Together with a battery management system (BMS) and safety switches, it makes up a usable battery system.

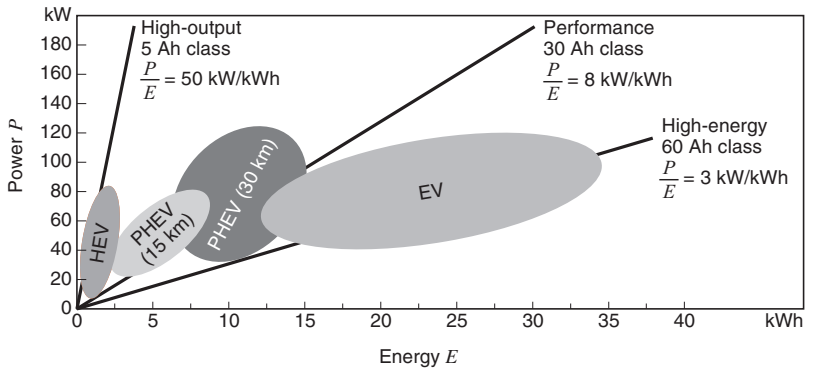
The (sometimes conflicting) requirements described below must be taken into account when developing and designing the cells and other components in the battery system.

Safety

The safety of vehicle occupants and the environment must be guaranteed in all situations. This applies to hazards that result from the electrochemical properties of the battery, for example, preventing fire. It must also take electrical safety into account, particularly in batteries with voltages larger than 60 V.

Figure 1: Requirements for energy content and capacity of the battery for various vehicle concepts

HEV Hybrid electric vehicle,
PHEV Plug-in hybrid electric vehicle,
EV Electric vehicle.



Performance
For use in vehicles, batteries are required to be low in weight while simultaneously having high performance. Achieving this requires batteries with high energy and power density. There are different requirements for this depending on the vehicle. Figure 1 shows the typical values for the respective characteristics and the ratio of performance P to energy E. This results in various requirements for the cells used. Various cell sizes and designs are required to be able to cover all vehicle segments.

Operation in the automotive sector requires a service life usually lasting more than ten years or a running performance of typically 250,000 km. In purely electric vehicles, approximately 3,000 complete charging cycles occur throughout the life of the vehicle, while applications in hybrid vehicles often have over a million partial charging cycles. The largest challenge by far is to reduce the costs with the goal of reaching total costs comparable to those in systems with internal-combustion engines.

Storage technologies

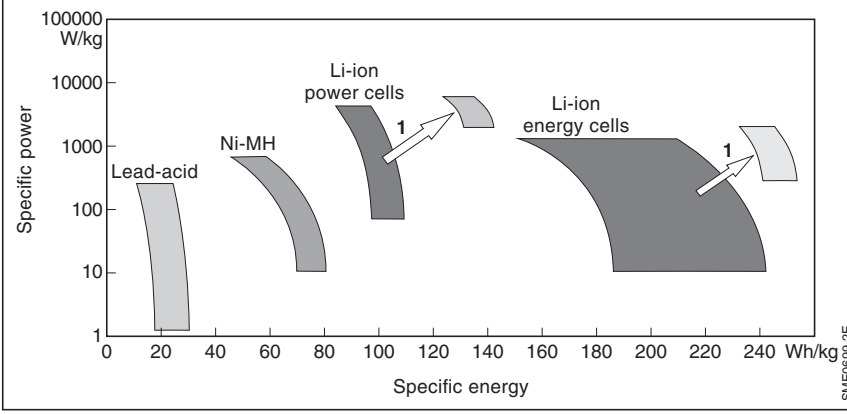
Today, lithium-ion (Li-Ion) batteries are usually used as electrical energy accumulators in electric vehicles, hybrid vehicles and plug-in hybrid vehicles. Nickel-metal hydride batteries (NiMH) are still used in isolated cases. However, it is to be expected that these will be gradually replaced by lithium-based battery systems.

Figure 2 shows a comparison of specific output and specific energy of various technologies.

Nickel-metal hydride technology
Electrochemical principle
In an NiMH cell (nickel-metal hydride), a metal alloy which can accumulate hydrogen is used as the anode [1, 2]. The active material of the cathode is nickel(II) hydroxide. During the battery-charging process active hydrogen is created from H_2O at the anode (Figure 3b). This hydrogen is absorbed by the metal alloy (M) of the anode and a metal hydride (MH) is created. The remaining OH^- ion migrates through the electrolyte and the ion-permeable separator to the cathode, where it reacts with the nickel(II) hydroxide to form NiO(OH).

During discharging the accumulated hydrogen at the anode oxidizes with the OH^- ions created at the cathode to pro-

Figure 2: Typical specific energies and capacities for various storage technologies (Ragone plot)
1 Development goal.



duce water (Figure 3a). The following reactions take place during the discharging process; these are reversed during the charging process.

Anode: $MH + OH^- \rightarrow M + H_2O + e^-$ (0.828 V)

Cathode: $NiO(OH) + H_2O + e^- \rightarrow Ni(OH)_2 + OH^-$ (0.450 V)

Redox equation: $MH + NiO(OH) \rightarrow Ni(OH)_2 + M$ (1.278 V)

Application

Nickel-metal hydride batteries have already been in use as traction batteries for a few years, in particular for hybrid vehicles. They deliver high currents and have a high charging capacity while simultaneously being low in weight. The primary disadvantage of cells in nickel-metal hydride technology, however, is the lower energy density at a specific power that is also usually lower compared to lithium-ion batteries. Nickel-metal hydride cells achieve a specific energy of up to 80 Wh/kg at a voltage of typically 1.25 V per cell. Nickel-metal hydride cells also have higher self-discharge than lithium-ion cells (approximately 1 % per day at room temperature), limiting their use to devices with short downtimes.

Lithium-ion technology

Electrochemical principle

The lithium-ion battery utilizes the reversible inclusion and withdrawal of lithium ions (Li^+) in a lattice (intercalation electrode, [1, 2]). Graphite (C) in various modifications, for example, is used as the anode material. The cathode is based on a structure of lithium transition metal oxides. The active material today consists for the most part of a combination of nickel, cobalt and manganese (Ni Co Mn), each constituent making up around a third of the overall compound. Lithium ions are exchanged during the charging and discharging process via the electrolyte between the anode and the cathode. The cathode and the anode are insulated against each other by the ion-permeable separator.

Figure 3: Electrochemical processes in a nickel-metal hydride battery
a) Discharging,
b) Charging.
1 Anode, 2 Cathode, 3 Alkaline electrolyte,
4 Separator, 5 Cell housing.
R Resistance, *I* Charging current.

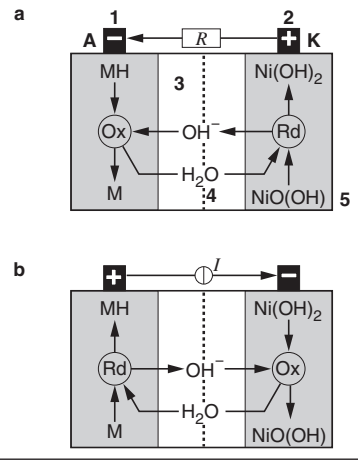
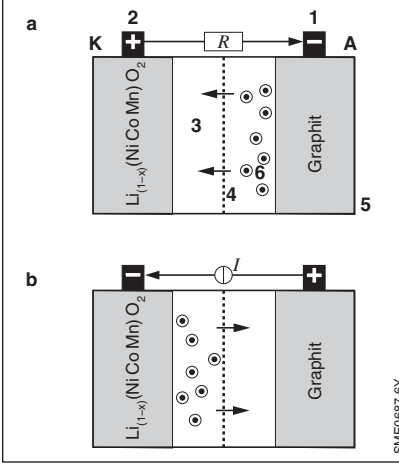


Figure 4: Electrochemical processes in a lithium-ion battery
a) Discharging,
b) Charging.
1 Anode, 2 Cathode, 3 Organic electrolyte,
4 Separator, 5 Cell housing, 6 Li^+.
R Resistance, *I* Charging current.



The following reactions occur (Figure 4):

Anode: $Li_xC_n \rightarrow n\,C + x\,Li^+ + x\,e^-$.

Cathode: $Li_{(1-x)}(NiCoMn)O_2 + x\,Li^+ + x\,e^- \rightarrow Li(NiCoMn)O_2$.

Balance equation: $Li_{1-x}(NiCoMn)O_2 + Li_xC_n \rightarrow Li(NiCoMn)O_2 + n\,C$.

Another established cathode material is lithium iron phosphate ($LiFePO_4$).

Depending on the anode used (graphite, soot, "coke carbons"), which is expressed by the index n, the proportion x of the lithium ions that can be intercalated changes. The electrolyte consists of anhydrous organic solvent mixtures (e.g., organic carbonates such as ethylene carbonate, dimethyl carbonate, diethyl carbonate, and ethyl methyl carbonate), various additives, and conducting salts (e.g., $LiPF_6$).

Application

The lithium-ion battery is characterized by a high energy density and thermal stability. The nominal voltage is around 3.7 V per cell, where the voltage limits during use are between around 2.8 V and 4.2 V.

The simultaneous optimization to a high output and energy density is a technical challenge. A high power density requires large electrode surfaces with low-resistance conductors, while a high energy density requires the opposite: compact electrodes with a large amount of active material. Typical characteristic values of today's automotive cells include 5,000 W/kg for the specific performance of power cells and 180 Wh/kg for the specific energy of energy cells. Lithium-ion systems with a higher specific energy are in development (see Figure 2).

System monitoring

Lithium-ion systems – compared to other technologies – require more elaborate system monitoring and control systems using a battery management system. A lithium-ion battery must be kept within strictly defined charge, current and temperature limits for service life and safety reasons. In particular, lithium-ion cells do not tolerate overcharging. A battery management system (BMS) is used to prevent this.

Service life

The service life of electrochemical cells depend in an existing chemical system on the following parameters: number of cycles, depth of discharge, temperature, and mechanical load. Above all processes such as the formation of coatings, irreversible compounds, cracking, and loosening of the active material due to changes in volume take place at the electrodes. These processes manifest themselves electrically in an increase in the internal resistance, which results in reduced power and a reduction of the energy content that can be drawn. The definition of end of life (EOL) in a specific application depends on the boundary conditions of that applications; these are primarily the demands imposed on the operation of vehicles.

Further developments

It is to be expected that cells with innovative separator systems or based on electrochemical approaches, such as lithium-sulfur or lithium-air, are able to provide significantly higher energy and power densities at equal or even lower costs. However, systems based on this are still being researched. The capability for mass production for use in the automotive industry has yet to be proven.

Basic structure of a battery system

Traction batteries also differ in their internal structure due to each of the different technical requirements and individual installation situations in the vehicle. However, most follow a similar basic concept, which is depicted below.

A battery system usually consists of multiple modules, each of which, in turn, contains multiple cells. The modules are connected to each other via a high-voltage wiring harness and are integrated together with the battery management system and the (air or liquid) cooling system in a battery housing. Interfaces for the vehicle include the high-voltage terminals of the battery, a data interface and usually an additional connection for the cooling system (Figure 5). The battery management system can be situated outside the battery housing.

Components of a lithium-ion battery system

Lithium-ion cells

Depending on the application, a drive battery consists of battery cells, ranging from ten to one hundred and, in individual cases, even thousands, which are connected in series or in parallel. Here, cells are used in three different designs (shown in Figure 6) for use in automotive industry.

Designs

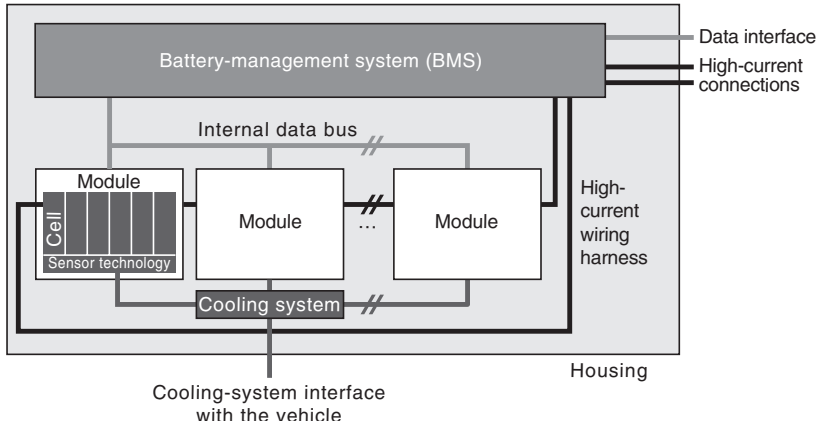
Cells with cylindrical metal housing have major cost advantages in manufacturing when compared to the other designs.

Prismatic cells with a housing made of laminated metal foil (pouch cells) have a particularly homogeneous temperature distribution and permit highly compact battery systems of flexible sizes.

Prismatic hardcase cells also permit a very compact construction and are also stable in comparison.

It is still uncertain whether one of the three cell types will prevail for automotive applications. Currently, there are examples of all three designs. There are global efforts to standardize the types of cell housing for automotive applications as well.

Figure 5: Schematic representation of the architecture of a battery system



Design of a lithium cell

Figure 7 shows the cross-section of a prismatic hardcase cell. The active materials required for the chemical process in the anode (e.g., graphite) and the cathode (lithiated metal oxide such as Li (Ni Co Mn) O_2 or lithiated metal phosphate such as Li Fe PO_4) are present as coating on a copper film (anode) and an aluminum film (cathode). The metal films are each used here as shunts. The films, coated on both sides, are separated by a separator film that prevents a direct electrical connection between the anode and cathode, but is permeable to ions. The ion-conductive electrolyte (organic liquids with fluoric conducting salts) is also located between the anode and cathode. This sequence of anode, separator and cathode is wound into a flat coil of cells, the electric shunts at the anode and cathode lead to the negative/positive electric terminal of the cell.

An additional safety valve is also located on the cell. If the pressure in the cell rises in the case of a malfunction, it opens to prevent the cell from bursting.

Safety requirements

When a technology-dependent temperature threshold in the cell (typically around 100 °C) is exceeded, a self-reinforcing, irreversible exothermic process called "Thermal Runaway" starts and can no longer be stopped by external influences. This can cause the cell to catch fire or burst and spread to other cells. Therefore, corresponding measures (voltage and current monitoring, cooling) must be taken to ensure that this process is not triggered by improper operation. This is the function of battery management.

Mechanical and electric components

Cells have to be interconnected in a suitable form and mechanically integrated for use in vehicles. This accounts for the mechanical and electrical safety requirements that are particularly applicable for batteries with high voltage. Typical automotive industry requirements also apply here, both in use (e.g., protection against vibration protection or splash water) and in potential cases of malfunction (e.g., protecting occupants and the environment in case of an accident).

Figure 6: Designs of lithium-ion cells

a) Cells with a cylindrical metal case,
b) Prismatic cells with case made of laminated metal foil (pouch cell),
c) Prismatic hardcase cell.

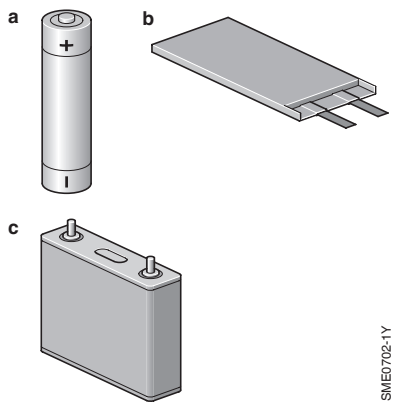
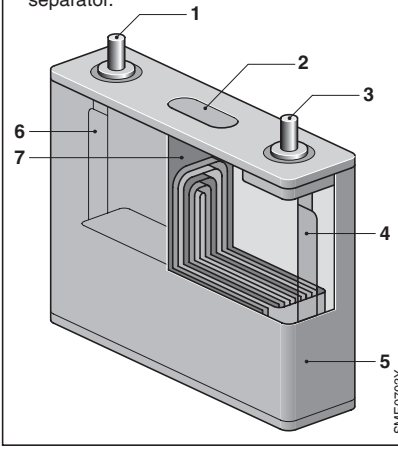


Figure 7: Cross-section of a prismatic hardcase cell

1 Positive terminal (pole),
2 Safety valve,
3 Negative terminal (pole),
4 Current collector,
5 Cell case,
6 Current collector,
7 Cell winding with anode, cathode and separator.



Module
The module (Figure 5) is usually the first integration level of cells for a battery pack. Here, a different number of cells is connected in series depending on the cell type. A parallel connection of cells makes it possible to deliver greater energy contents without increasing the system voltage. Usually, care has to be taken to ensure that a module voltage of 60 V is not exceeded in order to prevent high-voltage requirements at the module level.

The cells are mechanically integrated into a network and their terminals are connected by a high-voltage cable harness. Additionally, sensors for temperature, voltage and current measurement and evaluation electronics are frequently integrated into the module (control unit on the module level). The focus here is to ensure the most compact design engineering possible that allows flexible, cost-effective and high-quality battery production.

While small battery systems can only consist of one module, battery packs for electric vehicles often use over ten modules. The modules are connected to each other electrically via a high-voltage cable harness.

Housing
The battery housing, along with holding together the cells and protecting them from mechanical loads, is also used for contact protection against high voltages. On the outside of the battery housing, there are high-voltage connections for connecting to the electric drivetrain, which can be disconnected from the supply by a safety switch when the battery is removed, a communication interface for data exchange with the higher-level vehicle control unit for the electric drive (e.g., engine management system) and, in many places, interfaces for the cooling system.

Battery management system
Determining characteristics of the battery system, complying with the permissible operating limits and regulating the cooling system are necessary for safe and reliable battery operation. The battery management system (BMS) assumes these tasks. It is usually implemented in a control unit located inside the battery. However, individual functions can also be assumed by external control units.

Operation
The battery management system determines the instantaneous battery parameters and stores these if necessary to be able to run fault analysis or else for decision guidance for later applications outside the vehicle (e.g. “second-life applications” as stationary a storage device in solar installations). It uses the measured values to determine the state of charge of the cells (SOC) and their state of health (SOH). The currently available battery rating and maximum permissible battery current can be determined based on this. The battery management system can also carry out “balancing” to match the cells’ state of charge. This information is transmitted to the central vehicle control unit (VCU) in the vehicle's electric drivetrain. The vehicle control unit controls the power electronics in the electric drivetrain or the charging device based on the information sent by the battery management system to configure a corresponding charging process.

The battery management system protects the cells by preventing or reducing the current flow in cases when the battery is overcharged, the safe operating window is exited due to overvoltage or undervoltage, or in the event of excess or insufficient temperature. This occurs by instructing consumers to switch off the vehicle control unit in case of emergency by shutting down the battery on its own.

Even if the battery itself has to be assigned to the electrochemical domain, the functional safety standard (ISO 26262 [3]) for electrical and electronic components must be taken into account when developing the battery’s safety concept for the battery management system. Safety-related components of the battery management system, particularly the sensor technology, are, therefore, redundant.

Architecture
The architecture of a battery management system is illustrated by way of example in Figure 8. Other architecture variants are also used, particularly for hybrid electric vehicles (HEV).

The battery management system consists of a central battery control unit (BCU), cell supervisory circuits (CSCs), current, voltage and temperature sensors, and a contactor circuit. The cell supervisory circuits monitor the battery cells in terms of their voltage and temperature.

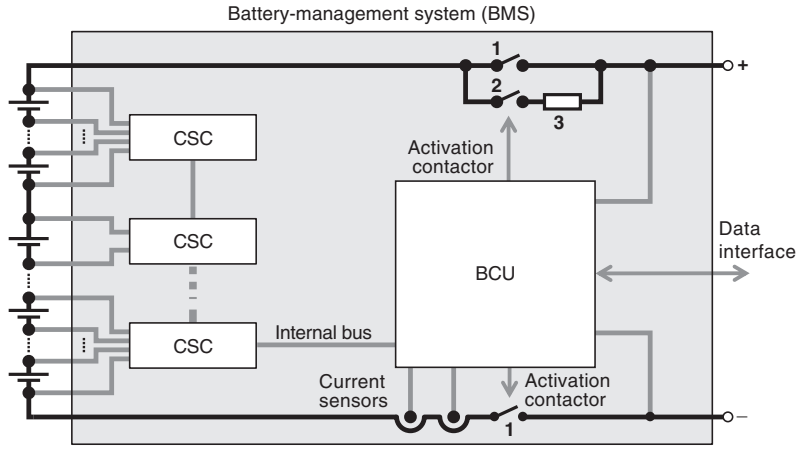
Conducting of current
The contactor current allows the vehicle electrical system battery of the electric drivetrain to be electrically isolated. If an accident is detected or another potential hazardous state is identified, a control logic opens the contactor that is located at each positive or negative battery terminal. To prevent a safety-critical condition, this logic can open the contactor even if the battery system is overloaded. The terminals of the battery system are thus de-energized.

The battery is connected to the vehicle electrical system of the electric drivetrain with its buffer capacitor via a charging contactor and a pre-charging resistor. This prevents an excessively large initial charge current when the buffer capacitor is discharged. During normal operation, the battery is connected to the vehicle electrical system with low impedance via the main contactor.

Balancing
Due to manufacturing tolerances and side reactions that depend on cell parameters and the cell temperature, the individual cells eventually have a different state of charge. This is problematic insofar as the cell with the lowest state of charge sets the discharge limit, and the cell with the highest state of charge sets the charge limit. For this reason, the state of charge of a battery's cells has to be balanced from time to time. This is done by a balancing circuit, which is usually integrated in the cell supervisory circuits. In automotive applications, balancing is typically a passive process; that is, it occurs by discharging individual cells via a resistor connected in parallel.

Thermal management
To ensure safe and reliable operation of the battery even when the load is fluctuating and to maximize the service life of the cells, the temperature of the cells should be kept in a range typically from 20 °C to 40 °C. The service life of the cells is largely

Figure 8: Overview of the battery system
1 Main contactor, 2 Charging contactor, 3 Pre-charging resistor.
CSC Cell Supervisory Circuits,
BCU Battery Control Unit.



reduced as the temperature increases, because many aging processes are temperature-dependent. This is particularly relevant for use in hot climate zones. At low temperatures, the capacity of the cells noticeably decreases.

In vehicles, there are various systems used to control the temperature of cells.

Liquid cooling
Liquid-cooling systems are usually used in systems with high requirements for regulating temperature. The battery is equipped with a separate cooling circuit, in which circulates a special coolant, such as a water-glycol mixture. The coolant temperature is controlled by a heat exchanger, which is connected to the vehicle's air-conditioning system. Frequently a heater element is also integrated, so that the battery can be heated at low temperatures.

It is connected to the cells via cooling plates. Temperature sensors monitor the temperature of the cells. The central thermal management system, often integrated in the battery management system, controls the flow rate and coolant temperature depending on the situation and temperature.

Batteries with a liquid-cooling system can be built very compactly, but require additional effort due to the cooling plates, the piping and the connection to the vehicle's air-conditioning system.

Refrigerant cooling
An alternative to a liquid-cooling system is the refrigerant cooling system, in which the cooling components of the battery are designed as evaporators and directly integrated into the refrigerant circuit of the vehicle's air-conditioning system. Systems of this kind have been used only rarely to date.

Air cooling
Air-cooled systems are the most cost-effective designs. However, they require more space due to the need for space between the cells. Moreover, the air interfaces mean that hermetically sealed battery cases are not possible, or they can be implemented only with great effort.

Air-cooled systems are often used for batteries with lower cooling requirements, for example, in systems with lower specific power requirements and for use in moderate climate zones.

In air-cooled systems, fans controlled by a thermal management system take in air that is usually precooled. This is blown through the battery system; the fan speed can be controlled depending on the cell temperature determined by the temperature sensors.

Some systems for low power requirements have no fans; in that case, the cells are simply cooled by natural convection.

Charging method for lithium-ion batteries

Lithium-ion batteries are generally charged according to the IU method, i.e., with constant current up to a specified voltage, then further with constant voltage at a specified maximum current. The battery can be charged relatively quickly with very good efficiency up to a state of charge of around 80 %; this is exploited by fast-charging methods. The limit values for current and voltage are specified by the cell manufacturer and vary depending on the electrochemical system used. Adherence to these limit values is monitored by the battery management system (BMS) because exceeding these limit values may give rise to safety-critical states.

Recycling of lithium-ion vehicle batteries

Lithium-ion vehicle batteries consist of a large number of cells and of housing and electrical and electronic components, and have significant economic value. To utilize these batteries to their full potential, scientists and engineers are currently examining the possibilities of their stationary use after automotive use. This is referred to as "second life". Applications may include for example the storage of solar power.

The batteries are subjected to a recycling process when they reach the end of their service life. For this purpose the batteries must be fully discharged for safety reasons and then mechanically disassembled to the cell level. The housing and the electrical and electronic components are processed further in accordance with tried-and-tested methods. The cells are further disassembled mechanically and later the substances of content are recovered by means of hydrometallurgical and pyrometallurgical processes. The aim is recover the substances of content of such quality as to permit their use in electrochemical cells. Processes are currently being tested to assess their economic viability.

References

[1] C.H. Hamann, W. Vielstich: Elektrochemie. 4th Ed., Wiley-VCH, 2005.

[2] A. Jossen, W. Weydanz: Moderne Akkumulatoren richtig einsetzen. 2nd Edition, Springer-Verlag, 2019.

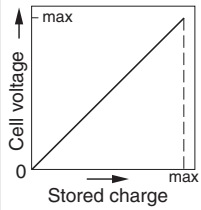
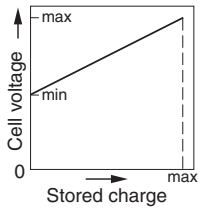
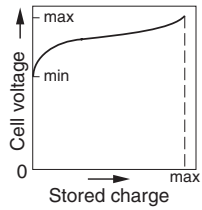
[3] ISO 26262: Road vehicles – Functional safety.

Supercapacitors

Application

Supercapacitors are classed among the rechargeable electric storage cells such as, for example, electrolytic capacitors and lithium-ion cells. In capacitors, electrical energy is stored in an electric field; in a battery, energy is stored through the process of an electrochemical reaction. To achieve a voltage that can be utilized for vehicles (e.g., 12 V for vehicle electrical systems or 400 V for electric drives), it is necessary to connect in series both supercapacitors and electrochemical storage cells since the individual cell voltages are restricted by the stability of the electrolytes and electrode materials used. The required energy for the electric storage devices is, taking account of the finally required voltage, then set by the capacitance of the individual cells and the number of cells connected in parallel.

Table 1: Comparison of storage principles

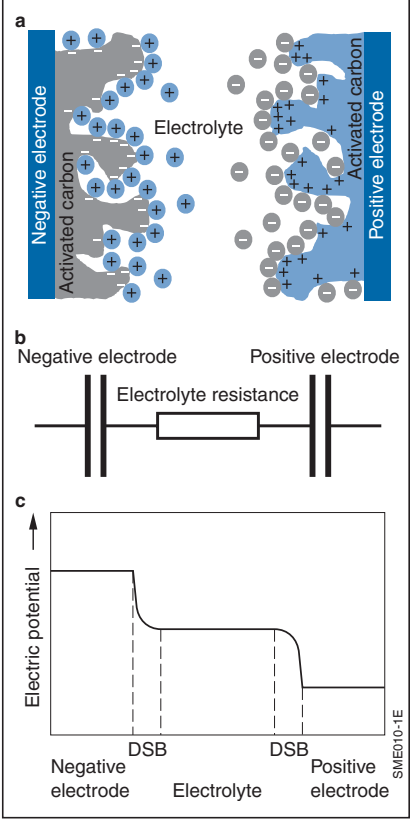
	Double-layer supercapacitor	**Hybrid supercapacitor**	**Power-optimized lithium-ion cell**
Reference technology: Negative electrode Electrolyte Positive electrode	 Activated carbon Organic electrolyte Activated carbon	 Graphite Organic electrolyte Activated carbon	 Graphite Organic electrolyte $Li(Ni_xCo_yMn_z)O_2$
Idealized steady-state voltage characteristic			 SME0709D
Typical working range [V]	0 → 2.7 to 3.0	2.0 → 3.8	2.5 → 4.2
Formula for energy content	$\frac{1}{2}CU_{max}^2$	$\frac{1}{2}C(U_{max}^2 - U_{min}^2)$	$\int_{Q=0}^{Q_{max}} U(Q)dQ$
Specific energy [Wh/kg]	4 to 8	10 to 20	80 to 120
Energy density [Wh//]	5 to 10	15 to 20	150 to 250
Specific power [kW/kg]	5 to 15	4 to 8	2 to 5
Number of full-charging cycles over service life	> 1,000,000	> 100,000	> 1,000

Electric double-layer capacitors

Storage principle

In supercapacitors of this type, energy is stored in the electric field of the electrochemical double layers of the negative and the positive electrodes. This creates opposing charges of different polarities in the area of a solid/liquid interface. There is surface enrichment and enrichment of electrons on the solid side, while positively and negatively charged ions are accumulated on the liquid-electrolyte side. There is no electrochemical reaction involved.

Figure 1: Double-layer capacitor
a) Real geometry,
b) Equivalent circuit diagram,
c) Potential profile.
DSB Double-layer range.



Electrode materials

Large surface-area carbon materials (especially activated carbon) are used as electrode materials [1], which are applied in an open-pored structure together with binders and conductive additives (e.g., soot) to metallic collectors. Different methods (e.g., rolling, pressing, spraying), with or without the use of solvents, may be used to coat the collector film. The positive and negative electrodes can be made of the same materials.

Properties

Different working-voltage ranges may result, depending on which electrode materials are used and whether aqueous or non-aqueous electrolytes are used. Thus, the maximum cell voltage for non-aqueous electrolytes can be up to 3 V (a high voltage is beneficial for a high energy density); for aqueous electrolytes roughly only half of this value is achieved. The significantly better conductivity of aqueous electrolyte is beneficial for the internal resistance and the specific power of the component.

The use of new electrode and electrolyte components as well as additives offers optimization potential for increasing energy densities (unit: Wh/*l*) and power densities (unit: W/*l*). The aim here is in particular to extend the utilizable voltage window and to reduce the internal resistance. However, scientists and engineers are working on increasing effective surface areas, which includes pore geometries adapted to ion radii.

Like lithium-ion cells, large-capacity supercapacitors are also available in the cylindrical, “pouch”, and prismatic formats (see Batteries for hybrid and electric drives and [2]). The inner cell structure can also take the form of a coil or stack here.

Electric double-layer capacitors are characterized by very high specific power which is available in a comparatively large temperature window (roughly –40 °C to 70 °C) depending on the current direction. Furthermore, on account of the absence of electrochemical reactions for the stor-

age principle in question, they have very long cyclical and calendar lifespans and can be exhaustively discharged down to a cell voltage of 0 V. On the flip side, compared with power-optimized electrochemical lithium- or nickel-based storage cells, they have an energy density (unit: Wh/*l*) that is lower by a factor of 10 to 30 (see Table 1).

With regard to their electrical management in the system network, it is often sufficient to monitor only the voltages and temperatures. In addition, passive balancing methods – based on parallel-connected voltage dividers – are often sufficient to achieve equal voltages in a series connection.

Areas of application

In fully and partially electrified drivetrains, supercapacitors can assume the role of a reversible, electric storage device with high power output, for example for efficient energy storage during recuperation and to provide assistance when accelerating. Power, energy content, and voltage must be configured in accordance with the load requirements and the power-electronics peripherals. Double-layer capacitors can also be used in vehicles to power electric actuators and starters. Current applications for road vehicles are centered on the 12 V to 48 V range.

Hybrid supercapacitors

The term hybrid in the context of electric storage devices refers to an electrical linking of different storage technologies. This linking can be achieved by an electrical connection at the cell level or pack level, but also inside a storage cell, whereby storage materials of different technologies (e.g., supercapacitor and lithium-ion battery) are combined.

A significant increase in the specific energies (unit: Wh/kg) and energy densities (unit: Wh/*l*) of supercapacitors is achieved through the application of the principle of hybridization already at the storage-cell level. Here, the purely electric storage principle of double-layer capacitors is complemented by energy storage by means of reversible, electrochemical, Faraday processes. Such processes are characterized by an electron transfer at the boundary layer from solid to liquid. Hybrid supercapacitors currently on the market for the most part use electrodes and electrode materials along similar lines to those used in lithium-ion cells (or nickel cells for aqueous electrolyte concepts). The hybridization of energy storage can be realized in the area of just one or also both storage-cell electrodes. Finally, the capacitances of the storage processes active at the capacitor's respective voltage operating point are added up. An exhaustive discharge down to 0 V is for the most part not possible with this type of supercapacitor.

References

[1] P. Kurzweil, O.K. Dietlmeier: Elektrochemische Speicher, 2nd Ed., Verlag Springer Vieweg, 2018.

[2] A. Jossen, W. Weydanz: Moderne Akkumulatoren richtig einsetzen, 2nd Ed., Cuvillier Verlag, 2019

Electronic control unit (ECU)

Functions

Architecture

Digital technology with programmable controllers has revolutionized automotive engineering. A whole range of functions can be realized with this technology. Many influencing factors are taken into account in the control of the systems in the vehicle. The ECU (Figure 1) receives the electrical signals from the sensors, evaluates them, and calculates the triggering signals for the actuators. The control program – the "software" – is stored in a memory (e.g., flash EPROM). A microcontroller runs the program. The ECU components are called "hardware".

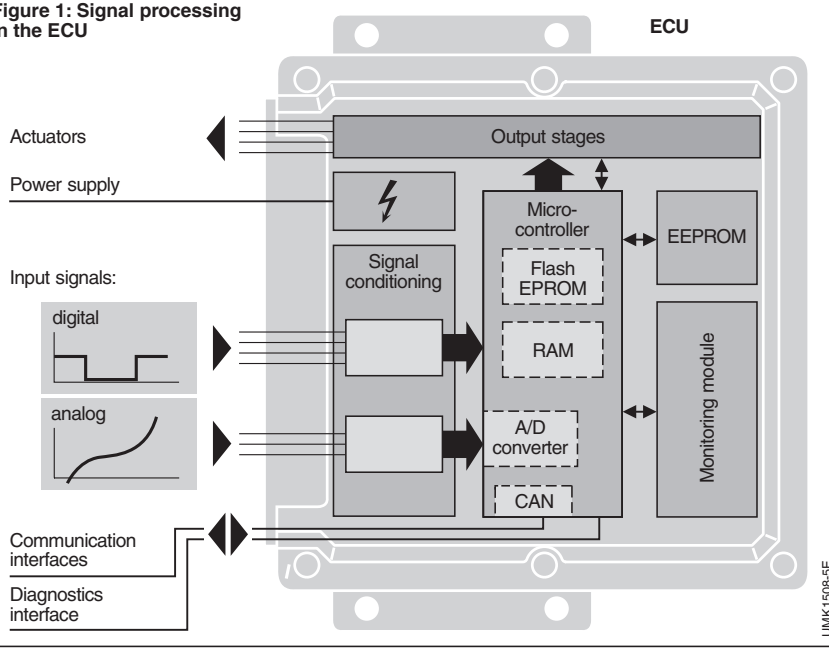
Examples of ECUs

Engine management, for example, performs the entire open- and closed-loop control of the engine (e.g., ignition in a gasoline engine and fuel injection) and of many assemblies in its periphery (e.g., for exhaust-gas turbocharging and exhaust-gas recirculation). Without these possibilities it would not be possible to adhere to current emission limits and low consumption figures under high engine-performance conditions.

Microcontrollers with high computing-capacity requirements are required for electronic open-loop control of systems. The electronic components are accommodated in an electronic control unit (ECU). The number of ECUs in a vehicle has increased sharply in the last few years; more and more systems are being electronically controlled. Further examples include:

– ECU for transmission control,
– ECU for driving-dynamics control (electronic stability program with antilock braking system and traction control system),
– vehicle power supply control unit,
– A/C control unit,
– door control unit.

Figure 1: Signal processing in the ECU



Compared with the ECU for driving-dynamics control for example the range of functions and the derived performance e.g. with regard to a door control unit, which controls the power-window drives, are low. Accordingly, less powerful components are used in the ECU for this purpose.

Electric and hybrid drives, which are increasingly penetrating the market, also have a large range of functions and thus impose high demands on the performance capability of open- and closed-loop control.

Operating principle
All ECUs operate according to the same principle. The ECU detects signals from sensors and control elements, evaluates them, and activates actuators (Figure 1). The ECUs are networked with each other via communication interfaces (e.g., CAN bus) so that they can exchange information with each other. An ECU network is thereby established. For example, the driving speed, which is required as the input signal in many ECUs, is only calculated in one ECU from the signals from the wheel-speed sensors and forwarded via the data bus to the other ECUs.

Future
Because of the performance capability of the electronic components in the ECU, systems can be combined in a single ECU. There are thus already systems in which engine-management and transmission control are combined in one ECU. The trend is towards “vehicle computers” (VCs) with high computing capacities and much more storage capacity which supplement the previous microcontroller-based ECUs. As a result, functions from conventional ECUs can be transferred to centralized “vehicle computers”.

Requirements

Operating conditions
High demands are made of ECUs. They are exposed to different and sometimes also extreme loads, depending on their installation location. Thus, for example, an engine ECU installed in the engine compartment or directly on the engine is subject to influences that are quite different from the door control unit installed in the passenger compartment. Irrespective of this, all ECUs must meet high requirements with regard to functionality, quality, and service life.

ECUs may be exposed to the following loads:
– ambient temperatures from –40 °C to +125 °C,
– temperature changes,
– vibration accelerations up to 20 g in a frequency range of 10 to 1000 Hz,
– effects of moisture and humidity, also from salt water,
– operating fluids such as oil, fuel and brake fluid.

The following additional demands are made of ECUs:
– operational reliability in the event of voltage fluctuations in the vehicle electrical system (e.g., voltage drop when cold starting),
– power dissipation up to 70 W,
– electromagnetic compatibility (resistance to electromagnetic radiation and low radiation of high-frequency interference signals),
– service life of 15 years for passenger cars, 30 years for commercial vehicles,
– active service life of at least 6,000 to 8,000 h for passenger cars, 30,000 to 50,000 h for commercial vehicles,
– kilometrage of 240,000 km (passenger cars) or 1,000,000 to 1,500,000 km (commercial vehicles).

Requirements imposed on ECUs are standardized in ISO 16750 [1]; there are also company standards for all automobile manufacturers.

Range of functions and data-processing speed

The performance of the electronic components used in an ECU is increasing continuously and facilitates ever more complex control algorithms. This applies particularly to the engine-management system, which must take ever more stringent emission limits into consideration. In response to new requirements the range of functions and thus the demand for memory capacity are increasing continuously, while the trend is toward ever-decreasing external device dimensions. Development is therefore heading towards greater functional integration and to miniaturization of both electronic and mechanical components such as plug-in connections.

The increasing range of functions is accompanied by an increase in data-processing speed. Functions must be executed within a specified time frame. Many program sections must be reliably executed with a predefined time period so that the processes in the ECU keep pace with the physical processes (in the sense of real-time capability).

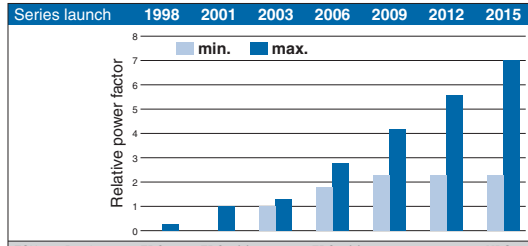
Development of complexity

The last few years have witnessed the use of ever more powerful microcontrollers with increasing computing speed and greater memory resources. At the beginning of the 1990s 32 kByte for the program memory, 8 kByte for the data memory and 12 MHz clock frequency were sufficient for an engine-management system. Microcontrollers currently used have program and data memories several MByte in size and several computer cores with several 100 MHz clock frequency.

The rising requirements imposed on ECUs (e.g., electronic engine management) and the growing number of functions are reflected in the computing capacity and in the memory capacity of the computers used. Conversely, it allows the advances in microelectronics made by information technology (IT) and the "consumer" sector to make this increased computing capacity available at relatively constant costs. Moore's Law [2] up to now also applies to automotive electronics (Figure 2). As a rule of thumb it can be assumed that the structure widths of the microcontrollers used in current engine control units will for around two to four generations (corresponding to around three to six years) will follow those of the current microprocessors. Currently (2021) structure widths of 7 nm are state of the art in the microprocessor field and 28 nm for engine control units in the microcontroller field.

And development continues apace, with microcontrollers with many computer cores being used in ECUs in the future (manycore computers).

Figure 2: Development of complexity of ECUs using the example of engine management for diesel engines [3]



ECU type Bosch	EDC15	EDC16(+)		EDC17(+)			MDG1
Word length, bit	16	32		32			32
Flash memory, MByte	0.3	1.5…4		1.5…6	2…8	1.5…8	1.5…8
Connector contacts	112…134	112…154		94…222		94…315	94…315
Clock frequency, MHz	20	40	56…66	80…150	80…180	80…260	80…300

SAE1378-1E

ECU components

The ECU is the control device of an electronic system. It detects the operating conditions via sensors, processes them and activates actuators. Signal processing takes place in the ECU's computer core.

Computer core

Microprocessor

The microprocessor (also MPU or microprocessor unit) represents the integration of a computer's central processing unit on a single chip. Microprocessor design ensures that units can be programmed to meet the varied requirements associated with specific operating conditions. There are two different main groups of processor. Earlier PCs (personal computers) used CISC processors (CISC: Complex Instruction Set Computing). A very large amount of special commands which are executed in a specific number of clock pulses is implemented in these processors.

RISC processors (RISC, Reduced Instruction Set Computing) are usually used in ECUs, i.e., also in vehicle ECUs. The advantage of RISC processors lies in the fact that the commands usually only require one clock pulse and can be implemented by the simpler and faster commands of higher-performance systems.

Today's RISC processors with more than 300 commands have more commands than older CISC processors, therefore the more accurate difference between CISC and RISC is the mean number of clock pulses for a command and no longer – as before – the number of commands.

An important variable for comparing the computing capacity of a computer architecture is the DMIPS specification (Dhrystone Million Instructions Per Second). It must be noted here that the computing capacity that can be achieved in current systems by system properties such as buses, technology and memory fluctuates by more than a factor of 5 (example: a 4 k DMIPS embedded MCU (microcontroller unit) needs 25 k DMIPS on an MPU). DMIPS is a very good specification for applications which require a very high bandwidth (predictable instructions and no interrupts). However, it is not suitable for embedded-control applications like engine-management systems in which the delay of the system in response to events in the time and angle ranges is important (e.g., crankshaft-angle sensing and angle calculation for injection, calculation and output of pulses for injection timing).

A microprocessor cannot usefully operate by itself; it always acts as part of a microcomputer.

Microcomputer

A microcomputer consists of a microprocessor serving as the central processing unit (CPU). The microprocessor contains the controller and the arithmetic-logic unit (Figure 3). The arithmetic-logic unit performs arithmetic and logic operations and operations involving digital signal processing, while the controller retrieves the commands and data from the memory. Pipeline stages, which preprocess the commands (e.g., retrieval of instructions and data, writing of data), are used to achieve higher clock frequencies. The arithmetic-logic unit is integrated in the pipeline.

The program memory (ROM, Read Only Memory; PROM, Programmable Read Only Memory; EPROM, Erasable Programmable Read Only Memory, or Flash EPROM) contains the non-volatile working program (user program) (see Program and data memory). The data are still available after switching on again even when there is no supply voltage. In current applications the program memory is usually a flash EPROM.

The data being run at any given time are stored in the data memory. These data change, and are stored in a RAM (Random Access Memory).

A cache as a high-speed buffer memory for program and data memories (in series production in Bosch engine-management systems since 2006) can be integrated on the microcomputer. If the required commands and data were always sent over buses and there were no cache, the high data rates (which the processor needs for working) and the relatively slow access to the memories would drastically reduce CPU speed and there would be no use in deploying a high-speed processor.

The bus system links the individual elements in a microcomputer. A clock generator makes sure that all the operations

in the microcomputer take place within a specified time frame.

Logic circuits are chips that carry out special tasks, such as for interrupts (interrupt) or the reset logic (to reset the microprocessor).

Input and outputs units (I/O) are required to connect with the peripherals. The peripherals include, for example, inputs for the crankshaft and camshaft signals, operating switches, analog inputs, and power drivers such as H-bridges (e.g., for throttle valve or exhaust-gas flap) or switch outputs (e.g., for heating of the λ sensor or relays).

Microcontroller unit (MCU)

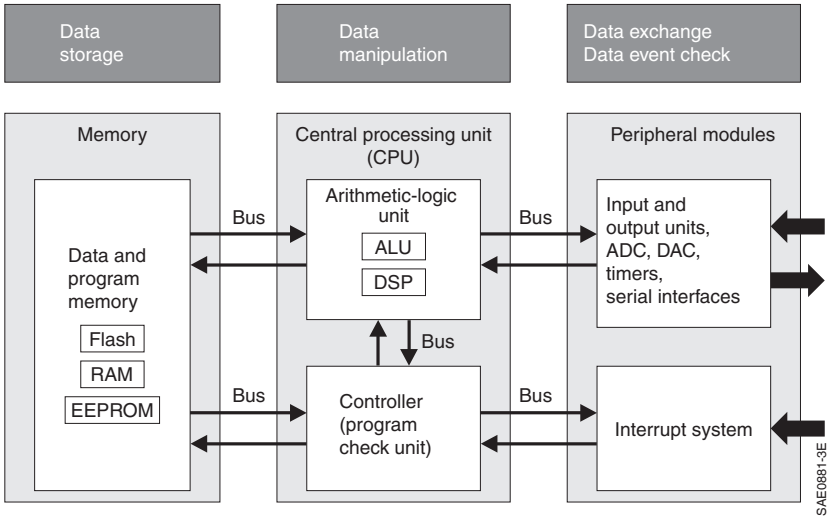
In-car applications are subject to the requirement that the sequence of the ECU program keep pace with the physical sequences. Thus, system must respond within the shortest possible time to, for example, changes in input signals. The systems must therefore have "real-time capability". Modern microcontroller unit have some circuit modules to facilitate this.

Computer core

The clock frequency for the computer core is up to 800 MHz. The computer core consists of up to eight cores which can alternatively be operated in lockstep mode. In other words, two cores receive the same data or instructions, the first core directly and the second core delayed by x clock pulses. The results of the first core are then temporarily stored for x (usually two) clock pulses so that they can be compared with the results of the second core (which come delayed by the x clock pulses). If the hardware is OK, the two results are identical. If not, there is a fault. In this way, ASIL C/D (classification regarding the criticality of the properties of a systems anticipated by the company – ISO standard for safety-relevant electrical and electronic systems in motor vehicles, ISO 26262, [3]) – can be supported.

Figure 3: Structure of a microcontroller

CPU Central Processing Unit, ALU Arithmetic-Logic Unit, DSP Digital Signal Processor, RAM Random Access Memory, EEPROM Electrically Erasable Programmable Read Only Memory, ADC Analog-Digital Converter, DAC Digital-Analog Converter.



Semiconductor memories
Semiconductor memories are required for the program code and for data (for details, see semiconductor memories).

DMA (Direct Memory Access)
This unit can transport data between the modules and thus relieves the CPU.

Interrupt controller
The interrupt controller signals to a computer core that, due to a time-critical event in the peripherals or in a different core, the processing of the current task may have to be interrupted.

Hardware accelerator
This is a module in which certain computing functions are implemented in hardware and not calculated by the computer core. The e-function and the MAC unit (Multiply Accumulate) are an example of this. MAC is a computing operation for the efficient calculation of mathematical expressions – two values are multiplied and added to the previous multiplications. This makes it possible today for example to process Gaussian functions by more than a factor of 50 faster than in a computer core based on functions programmed in a programming language. These accelerators will be necessary in future MCUs because a clock-frequency increase to provide this necessary computing capacity is neither technically possible nor commercially sensible.

GPIO
"General Purpose Input Output" switch inputs and switch outputs are read in and read out within 10 to 20 ms. Consequently this means no real-time capability.

Analog-digital converters
Analog-digital converters (AD converters) generated from an analog voltage signal (e.g., voltage generated by a thermistor located in a voltage divider) a digital value which can be processed in the processor. There are two basic principles here:
– Converter with a high absolute accuracy: The SH converter (Sample and Hold) stores the input voltage in a capacitor and then converts it with different processes into a digital value.
– Converter which continuously provides values with a high resolution: In the case of an SD converter (Sigma-Delta) only the delta for the previous measurement in output positively or negatively with a preset clearance as bit 0 or 1 for further processing in the module internally.

GTM (General Timer Module)
The function of the timer unit (today mostly GTM for embedded systems) is to encapsulate the real time. There are two basic principles here – capture and compare.

The capture function for input signals assigns a time or angle stamp (angle as reference for example to the crankshaft) to the events at the input and stores the capture values in registers. Thus the processor does not have to process the signal further immediately; it can first close the running software routine (e.g., a highly prioritized interrupt routine) and then process the capture values.

The compare function for output signals generates an event (e.g., switching of the ignition coil at the moment of ignition) at the output pin at a specific time or angle. These values are stored beforehand in registers.

Microcontrollers
Microcomputers which have real-time capability are called microcontrollers. Nowadays they are typically integrated on a chip and are not, as in the past, designed as discrete components on the printed-circuit board. This is possible due to the fact that highly complex systems can be integrated on a chip thanks to modern technologies.

For future applications it will no longer be sufficient to have just one CPU, since the clock frequency cannot be increased in any way the user likes and this also results in very high current inputs. For this reason, microcontrollers will come in variants containing more than one CPU. For series 2023 Bosch offers a system with 35 computer cores – 10 main cores, 12 digital signal processor cores, 10 cores in the timer, two cores in the peripherals, and one core in the security module.

The microcontrollers which are used in the automotive industry differ from standard microcontrollers on various points.
– Temperature range: Junction temperature from –40 to 165 °C.
– Failure rate: Less than 1 ppm (one failure per one million components) – compared with 100 to 200 ppm for PCs.
– Service life: This is current specified at 40,000 hours in active operation (this corresponds in the case of a PC that is switched on for eight hours a day to an operating period of approx. 14 years).
– Significantly longer availability: Whereas no new microcontrollers have to be subsequently delivered for cellular phones and PCs after a production cycle of 1 – 3 years, complex projects in the automotive field start series production only after 3 – 5 years development, after which they remain in series production for a period of 15 years and are then subsequently delivered for a further 15 years.

Types of microprocessor unit
There is no clear definition which determines whether a component is termed an MPU or an MCU on the basis of its properties. A sensible technical distinction is to make a distinction based on the type of application.

MPU
Such a computer core is optimized for the support of non-real-time functions like virtual memory management and the use of other virtualization mechanisms like Hypervisor to support operating systems like Adaptive Autosar or Linux. With such MPUs interfaces such as PCI (Peripheral Component Interconnect, a bus system for connection peripherals such as for example Ethernet switches and a computer core) or USB (Universal Serial Bus, a bus system for connecting different devices from the segment of entertainment electronics) are very often supported. The architecture is optimized for bandwidth – in this way plannable tasks can be processed at very high speed.

MCU
The microcontroller unit is designed in such a way that it can satisfy “hard real-time requirements” (reaction to input signals or setting of output signals in the µs range). Properties like virtual memory management and fast reaction times to external events are absent (because the applications in part exclude themselves – high frequency and virtualization as for MPU with fast reaction to events as for MCU). Because this architecture is optimized for short latencies, very large memories close to the core are required, and not externally as for the MPU.

Semiconductor memories

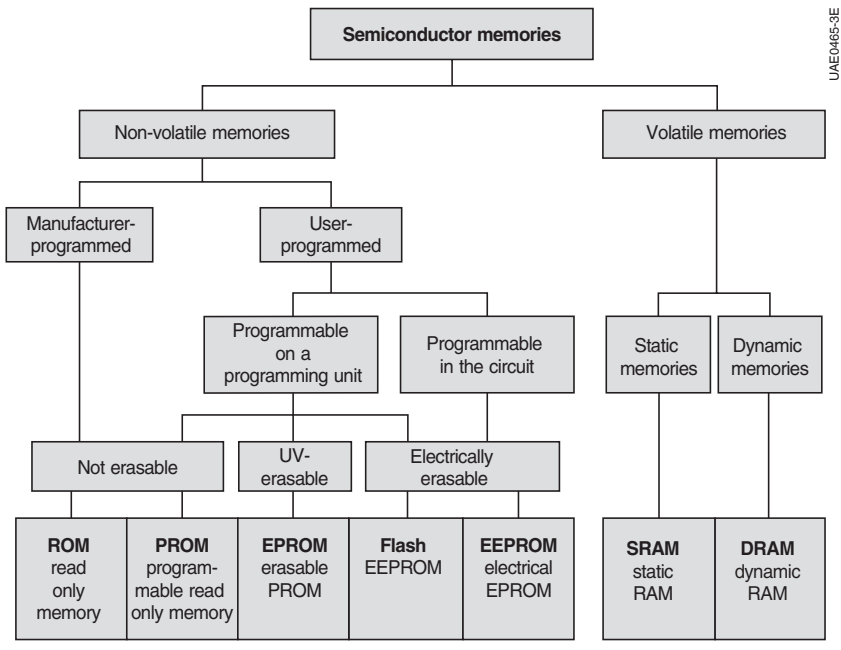
Memory principle

Data storage comprises the following operations: recording (writing), permanent storage (data storage in the narrow sense), retrieval, and output (reading) of information. Figure 4 provides an overview of the different memory types. The memory operates by exploiting physical properties that facilitate unambiguous production and recognition of two opposed states (binary information). In semiconductor memories the states produced are either "more conductive" and "less conductive" or "charged" and "discharged". A technology in frequent use today is the flash memory. Memories operating according to this principle are electrically programmable and erasable.

New memory types will also be used in the future. The FRAM (Ferroelectric Random Access Memory) with the ferroelectric memory principle, the MRAM (Magnetoresistive Random Access Memory) with the magnetic effect as the memory principle, and the PCM (Phase Change Memory) utilizes as its memory effect the change of state of a material from crystalline to amorphous and the associated change in resistance. MRAMs and PCMs have the potential to supersede the flash memories (non-volatile but slow) and RAMs (fast but volatile) used today, since they retain their data after the power supply is switched off and also have very short access times (to be introduced in engine management from 2023 onwards).

A major advantage of MRAMs and PCMs is that the bits can be individually reprogrammed and the cells do not have to be jointly erased as with the flash EPROM. This facilitates faster programming, which is advantageous for field updates outside the workshop and in the manufacturing process at the vehicle manufacturer. Another advantage is that the reprogramming of individual bits is possible in the application. Because the running times of

Figure 4: Overview of semiconductor memories



the devices is getting ever longer and the technologies are getting ever smaller, the memory cells contain much less charge and therefore individual bits could lose the stored information in the event of long running times and high temperatures. In modern microcontrollers with structures of under 90 nm there are two remedial measures to counter this phenomenon:

– Reprogramming: Mechanism to ensure that cells which lose charge have sufficient charge again – the storage cell is supplied with additional charge.
– ECC (Error Correction Code): Unlike reprogramming, in this case additional bits are stored in the memory matrix and erroneous bits are corrected with the information from these bits – customary here are two bit error detection and single bit error correction for 64 bit memory areas via additional bits (customary as today for flash and RAM in technologies under 90 nm) to correct the erroneous bits.

Only static memories (SRAMs) are used for ECUs which operate at internal temperatures of more than 110 °C. Unlike dynamic memories (DRAMs), SRAMs do not lose charge and do not have to be refreshed during operation in the ms time range. SRAMs keep their charge as long as the storage cell is connected to the supply voltage.

Program and data memory

The microcontroller requires a program – the software – for the calculations. It is stored in a non-volatile program memory. The CPU reads out the values, interprets them as commands, and executes these commands in order.

Variant-specific data (individual data, characteristic curves and program maps) are additionally stored in this memory. The software can be adapted to different vehicle variants (e.g., different ignition maps) with these data.

The software requires a read/write memory to store variable data (variables) such as for example computing values.

Flash memory

The flash EPROM has superseded the conventional EPROM erasable with UV light (Erasable Programmable Read Only Memory) as the program memory. It is electrically erasable so that the ECU can be reprogrammed in the service workshop without it having to be opened. The ECU is connected via a serial interface to the reprogramming station.

RAM

All the variable data calculated in the software are stored in the RAM (Random Access Memory, i.e., in a read/write memory). Modern 40 nm microcontrollers contain up to eight MByte RAM – therefore there is no need for an external memory module. Newer technologies will additionally feature a HyperRAM interface to make an external RAM available for the application.

The RAM loses the entire data stock when the ECU is disconnected from the supply voltage; it is therefore a volatile memory. Parts of the RAM can be permanently supplied with voltage to prevent adaptation values learned during vehicle operation from being lost when the ignition is switched off. These values are however lost when the battery is disconnected. Therefore this option is only very rarely used today and the adaptation values are written to the non-volatile EEPROM.

EEPROM

Data that change in vehicle operation and which must not be lost when the battery is disconnected (e.g., important adaptation values, codes for the immobilizer) must be stored in a non-volatile memory (e.g., EEPROM).

It is also possible to use separately erasable areas of the flash EPROM as a non-volatile data memory for these data. An EEPROM is emulated in the flash (in separate areas – outside the program and data memory) for this purpose. The current data in each case are written to the end of a table, which gradually fills the blocks of the flash area. The old data can then be completely erased with the block. This ensures that the correct data are in the memory even when the supply voltage is disconnected.

Monitoring module
A monitoring module is required for safety-relevant systems. This is effected by a logic circuit in the voltage regulator (e.g., for engine-management systems) or in a separate integrated circuit (e.g., for transmission management). The monitoring module can also be effected by a separate computer.

The microcontroller and the monitoring module use a "question-and-answer game" to monitor each other. When a fault is detected, the ECU of both is rendered safe independently of each other – in the case of engine management for example by the deactivation of torque-relevant output stages (examples: ignition and fuel injection; deactivating these outputs means that the engine cannot provide any power).

Voltage regulator
The electronic components in the ECU need a stable voltage of 5 V. Some applications also require further voltage values, e.g., 3.3 V. The voltage regulator maintains the battery voltage, which can be 6 to 16 V depending on the battery state and load, at these constant values. Suppressor circuits suppress high interference voltages from the vehicle electrical system.

The voltage regulator contains an enabling logic which ensures that the voltages run up under defined conditions and then enable resets. This ensures that the microcontroller runs up under defined conditions after the supply voltage is switched on.

ASIC chips
ASIC chips are application-specific integrated circuits. They are designed and produced in accordance with the specification of ECU development. Integrating a large number of functions in an IC reduces the space requirement, reduces manufacturing costs and increases operational reliability.

Thus, for example, the power supply for the ECU and several power output stages with the associated diagnostic circuits can be integrated on one chip. An example of this is the "U-chip", which features on one IC a power supply, the monitoring module, serial interface drivers and output stages, and an H-bridge with diagnostics.

Another function of an ASIC module is, for example, to relieve the load on the microcontroller or to make additional hardware available (e.g., additional input and output channels for static signals and for PWM signals).

An ASIC can be used in the gasoline-engine-management system for knock control which boosts the signals supplied by the knock sensors, filters them via a bandpass, rectifies them, and integrates them via a specified crank-angle window. The microcontroller then only has to read out the integrated voltage value at the end of the window. However, in newer ECU generations the knock-sensor signals are detected directly by the microcontroller via high-speed analog-digital converters (ADCs).

Special ASICs are also used to operate, evaluate and diagnose λ sensors. They activate the sensors with defined signals, evaluate currents, voltages and temperature with great precision, and convert these signals for the microcontroller into digital input signals.

ASICs are also used to activate fuel injectors for high-pressure fuel injection. A defined current curve can thus be set.

For special applications ASICs with several output stages and H-bridges are available which also contain a diagnosis function and can be activated via a serial bus.

Communication
ECU-internal communication with the ASICs is effected via SPI (Serial Peripheral Interface) or via MSC (Micro Second Channel).

Data which do not require a high bit rate (e.g., fault-memory data stored in the EEPROM) are also serially transferred over only one data line.

In this way, the ECUs can also send data for calculations (e.g., wheel speed, steering angle, battery voltage, drive torque, selected gear, engine speed) to other ECUs if they are connected to these via a multitude of communication buses.

The buses which are mostly used in vehicles are:
– CAN (with the subvariants CAN, CAN_FD and CAN_XL): for some devices up to 20 CAN buses to one ECU – data transfer currently at 1 MBaud, in future up to 15 MBaud.
– LIN: up to 18 LIN buses for the connection of sensors with low bandwidths.
– FlexRay: currently at 10 MBaud, will be replaced in future by a special Ethernet variant.
– Ethernet: at various speeds also up to 2.5 GBaud bandwidth.
– PSI5: for sensors also with return channel for avoiding sensor tuning.

In the case of tuning often the weakest link of a chain is attacked – if the engine control unit is more difficulty to tune, the modification of input signals is used to achieve higher engine power. 30 years ago the crankshaft-signal connector was connected with a 180 ° offset to achieve a more retarded ignition with regard to TDC (cylinder top dead center) and thus more engine power – with the drawback of influencing the engine's service life. Today, microcontrollers are switched into the signal lines, which relay to the engine control unit for example a driver's request for more power (via pedal-travel sensor or air mass) according to the table stored in the microcontroller – e.g., 10 % more with the same input data.

Data processing

Architecture

The engine control unit operates according to the classic IPO principle (Input – Processing – Output). It can be subdivided into three main blocks (Figure 5):
– input circuits for sensor-signal acquisition,
– computer for calculating the actuating signals,
– output stages with the power electronics for activating the actuators.

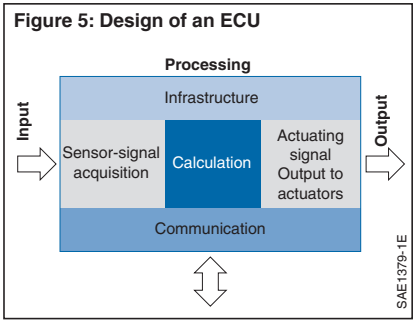
Sensors and actuators form as the peripherals the interface between the vehicle and the ECU as processing unit. In addition, there are bidirectional communication data buses (e.g., CAN bus), via which information is exchanged with other ECUs or sensors [4].

Input signals

Control elements and sensors form an interface between the vehicle and the ECU. The electrical signals are supplied to the ECU via the wiring harness and connectors. Sensors can have analog or digital interfaces.

Analog interface
Analog input signals can assume any value (e.g., voltage value) within a certain range. Examples of physical variables which are available as analog measured values are the coolant and intake-air temperatures, the intake-manifold pressure, the intake air mass, the battery voltage, and also the hydraulic pressure in the braking system. The analog voltage output by the sensors is relayed to the ECU. The analog value is converted by an analog-digital converter (ADC) in the microcontroller or in an interface module into a digital value. A typical resolution of the converters integrated in the microcontroller is 10 bits. For a reference voltage of 5 V a resolution of approximately 5 mV is obtained with these 2^{10} = 1,024 stages.

Figure 5: Design of an ECU



Digital interface
Digital input signals have just two states "High" (logic 1) and "Low" (logic 0). They can be evaluated directly by the ECU. Examples of digital signals are switching signals of control elements, speed signals from a Hall-effect sensor (for measuring wheel speed or engine speed), or static signals from a Hall-effect sensor (e.g., position sensor in the transmission).

Sensor signals via communication interface
In future sensors which detect analog quantities will increasingly incorporate additional electronics which digitize the analog voltage locally and output the measured value via a standardized digital interface (see Integrated sensors). Examples of these interfaces are SENT (Single Edge Nibble Transmission) and LIN (Local Interconnect Network), whereby data can be transmitted bidirectionally (see LIN). They enable sensors from different manufacturers to be connected to ECUs.

Signal conditioning

Adapter circuits
Each sensor has a specific input. The input signals are limited by suppressor circuits to permissible voltage levels. Filters remove most of the superimposed noise from the useful signals and limit the bandwidth (sampling theorem); the useful signals are then amplified or attentuated if necessary to the microcontroller's permissible input voltage (0 to 5 V).

Signal conditioning in the sensor
Depending on the sensor's integration level, signal conditioning can take place in part or entirely in the sensor already. The sensor signals digitized in the evaluation electronics are then output via a communication interface; further special signal conditioning in the ECU is not necessary.

Signal conditioning with ASICs
For some sensors, e.g., λ sensors, special modules (ASICs) are required which assume the complex tasks of activating and evaluating these sensors. These ASICs activate the sensors with defined signals, evaluate currents, voltages and temperature with great precision, and convert these signals for the microcontroller into digital input signals. The entire signal path (Figure 6) is achieved with a single ASIC.

Flexible input module
The market in the passenger-car sector is characterized by an increasing number of variants and niche applications to round off the manufacturers' product lines (for instance convertibles, roadsters, etc.). In the commercial-vehicle sector there has always been a very large number of special applications, for example bituminous finishers, complete beet harvesters, and other special machines. These often have special sensors; however, owing to the low quantities involved, it is not economical to create and maintain hardware variants. The trend is therefore towards software-configurable input circuits (Figure 7). The module L9966 used by Bosch (FlexIn ST Microelectronics) has for example 15 freely configurable inputs, which can be configured as analog input, resistance-measurement input, digital input, SENT (bus interface) or input for switching-type λ sensors. Furthermore, switchable current sources can be configured as pull-up or pull-down for each input.

Figure 6: Signal path for evaluating λ sensors

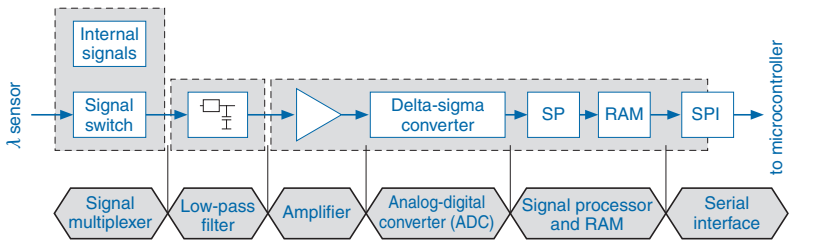
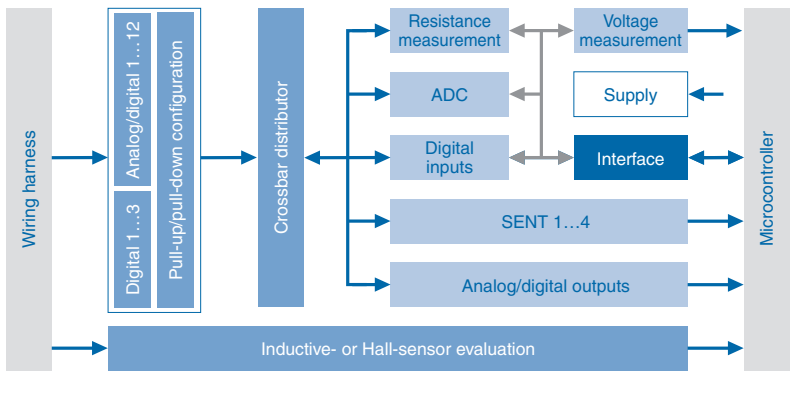


Figure 7: Block diagram of a software-configurable input circuit



Signal processing
The ECU is the control center for the functional sequences of an electronic system. The control algorithms run in the microcontroller. The input signals provided by the sensors and the interfaces to other systems for example via the CAN bus serve as input variables. They are checked for plausibility again in the microcontroller. The algorithms are processed with the aid of the ECU program (software). The ECU program is stored in the read-only memory (e.g., flash EPROM), the calculated intermediate values are stored in the data memory (RAM). The software calculates the output signals to activate the actuators.

Output signals
Different actuators are activated by the ECU via appropriate output stages on the basis of the data processed in the microcontroller. These actuators differ in their power demand and in how they are activated.

The output-stage modules in the ECU have a number of individually activated output stages which supply sufficient power to activate the actuators.

Signal types
Switching signals
Switching signals are used to switch actuators on and off, depending on the current operating point (e.g., fuel injectors, ignition coils). Consumers with high currents to be switched (e.g., engine fans) are switched via relays or with output stages directly on the assembly.

PWM signals
Digital output signals can be output as a PWM signal (pulse width modulation) with constant frequency and variable on-time. Actuators such as the exhaust-gas recirculation valve, the throttle valve or the charge-pressure actuator can, depending on this variable on-time, be moved into specified operating positions.

Application-specific output stages
Application-specific output stages are then used if special current and voltage curves for activating the actuators are needed which cannot be realized with standard output stages.

An example of such is the activation of high-pressure fuel injectors for gasoline direct injection (see High-pressure fuel injector) or for common-rail injectors (see Injectors). The actuators are activated by complex output-stage modules with integrated control logic and analog signal conditioning. These modules deliver a high voltage during switch-on, producing a high making current and thus a fast breaking operation. The system is then regulated back to a lower holding current.

Flexible programmable output module
Increased flexibility is also called for in the area of output drivers in order to cover different applications with one hardware variant. One solution is the flexible programmable output module DGDI-S from Bosch (originally for Diesel-Gasoline-Direct-Injection Solenoid, but not also for other applications, Figure 8). This module has four microprogrammable sequencers, seven high-side and eight low-side FET drivers, and six current measurements. It can be reconfigured by software for use for example for injectors, solenoid valves, H-bridges or BLDC motors (brushless DC motors). Consequently, it is no longer necessary to have all the inputs and outputs permanently defined already at the start of ECU development. Adaptation to the concrete application can be performed later by means of a software or data-record modification.

Switching variants
There are different switching variants for activating the actuators. Low-side output stages control inductive and ohmic loads which are connected to the battery voltage (e.g., valves, relays, ignition coils). High-side output stages switch consumers which are connected to ground. DC motors are activated via bridge-type output stages (Figure 9); here, both connections of the motor are switched by the ECU. This makes it possible to reverse the direction of rotation (e.g., for power-window drives, activation of the throttle valve).

Self-protection function
The output stages are protected against short-circuits to ground, against short-circuits to battery voltage, and against being destroyed due to electrical or thermal overload. These malfunctions and interrupted lines are detected as faults by the output-stage IC and signaled to the microcontroller via a serial interface.

Communication interfaces
The ECU has one or more communication interfaces for communicating with other electronic systems. In this way, for example, the driving speed can be sent by driving-dynamics control to all the systems that need this signal (e.g., the instrument cluster for displaying the driving speed). The signals therefore only have to be calculated one in the ECU network. Data transfer is effected via bus systems (see Automotive networking, buses in the vehicle).

Figure 9: H-bridge for activating a DC motor

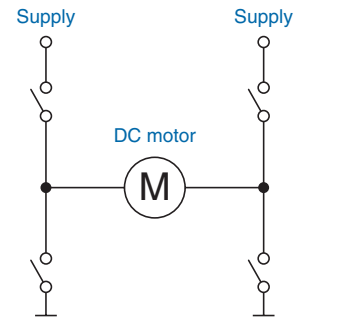
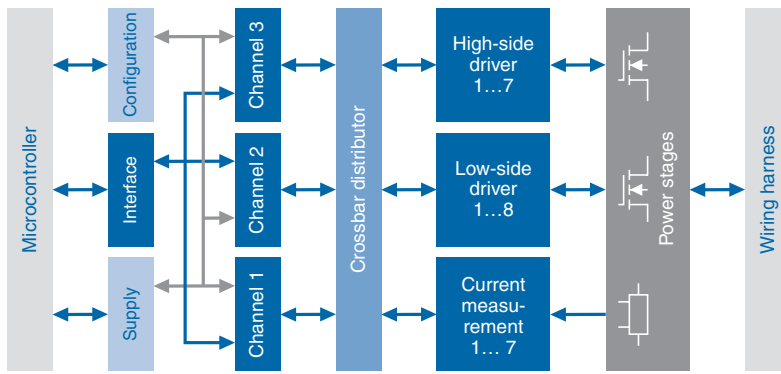


Figure 8: Flexible programmable output module DGDI-S



Assembly and connection technology

Circuit-board technology

Engine control units (Figure 10) are typically manufactured in accordance with circuit-board technology using circuit-board materials with four to eight layers. The housing consists for example of a base/cover combination of diecast aluminum and deep-drawn aluminum sheet which are screwed or bonded to each other. A membrane impervious to water is used for pressure equalization. As a result, an atmospheric-pressure sensor can be installed in the ECU which for engine-management purposes enables the injected fuel quantity also to be corrected on an air-pressure-dependent, i.e., altitude-dependent, basis.

Hybrid control units

In special cases, especially where there are extreme temperature requirements, thick-film hybrid circuit-boards are used. This technology involves the application of both integrated and discrete components on ceramic substrates.

Plug-in connections

A customary plug-in connection for example for engine management in passenger-car applications has around 200 contacts, split into a number of compartments. In commercial-vehicle applications a number of connectors are also placed next to each other on the cover, making possible many more contacts (Figure 10).

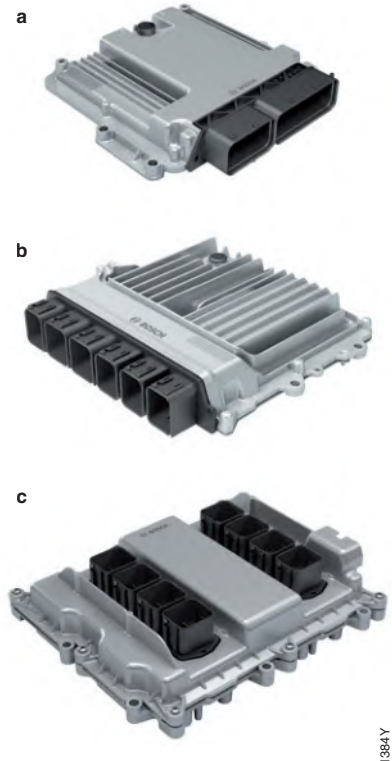
Small ECUs (e.g., a sensor control unit) as a rule have one plug-in connection with one compartment and few contacts.

Cooling

Heat dissipation is particularly important for the electronics in a vehicle's engine compartment (e.g., engine control unit, driving-dynamics control ECU). At ambient temperatures of up to 125 °C in the engine compartment the ECU's power loss, which can range from 30 to 70 W, must be dissipated outwards. There are

Figure 10: Examples of ECUs for diesel-engine control systems

a) Passenger-car ECU for body mounting (154 plug-in contacts),
b) ECU for passenger cars and light commercial vehicles (228 plug-in contacts),
c) Commercial-vehicle ECU for engine mounting for up to eight cylinders, diesel or natural-gas operation (336 plug-in contacts).



for this purpose special means of discharging heat from its point of origin outwards.

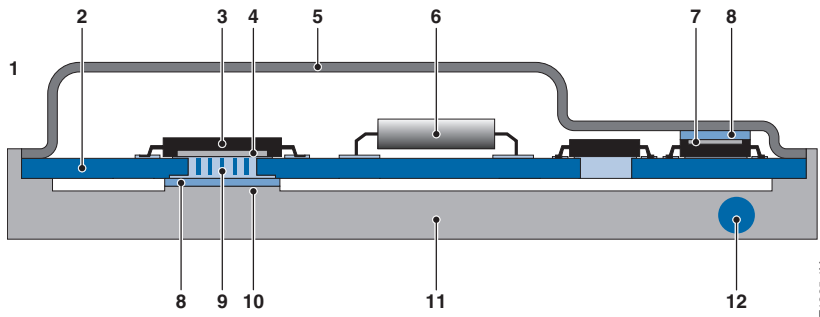
The left side of Figure 11 depicts the cooling of an IC (integrated circuit). Heat is dissipated to the printed-circuited board by the cooling surface situated below the IC (slug-down). Feedthroughs, which direct the heat to the underside of the printed-circuit board, pass through the printed-circuit board. The housing base is raised at this point so that the distance to the printed-circuit board is so small that it can be bridged with heat transfer compound.

The far right side of Figure 11 depicts the converse path. A component dissipates its heat via the cooling surface resting on it (slug-up) and the heat transfer compound upwards to the housing cover. It is shaped in such a way that the distance to the component is as small as possible. Heat dissipation from the housing to the surrounding area can be effected by way of heat radiation, heat conduction, and active heat transfer (forced convection).

Heat dissipation via the surface can be additionally boosted by cooling fins on the surface and additional movement of the air (e.g., with an external fan). To provide cooling by heat conduction, the ECU is mounted on a thermal conductor with as wide an area as possible (e.g., vehicle body).

Active heat transfer means that cooling ducts through which a cooling medium (e.g., fuel) flows are integrated in the ECU housing. Aside from the physical properties of the cooling medium, the magnitude of the heat flow is dependent on the coolant temperature, the flow rate, and the duct length.

Figure 11: Cooling of engine control units
1 Ambient temperature, 2 Printed-circuit board, 3 Power IC,
4 Slug-down (cooling surface under the IC), 5 Housing cover,
6 Capacitor, 7 Slug-up (cooling surface on top), 8 Heat transfer medium,
9 Feedthrough, in the p.c.b. under slug-down components for improved heat conduction through the p.c.b., 10 Cooling bank, 11 Housing base, 12 Fuel duct for cooling the ECU.



Software

To be able to support different electric and electronic architectures that are in a state of transition, it is necessary to isolate the application software from the hardware as far as possible. The solution comes in the form of standardized basic software (BSW) and standard interfaces for all the electronic vehicle functions. Generally, the basic software is defined as the hardware-dependent software, i.e., software components for example for microcontroller peripherals, drivers for application interfaces, diagnostics, communication, operating system, and system control.

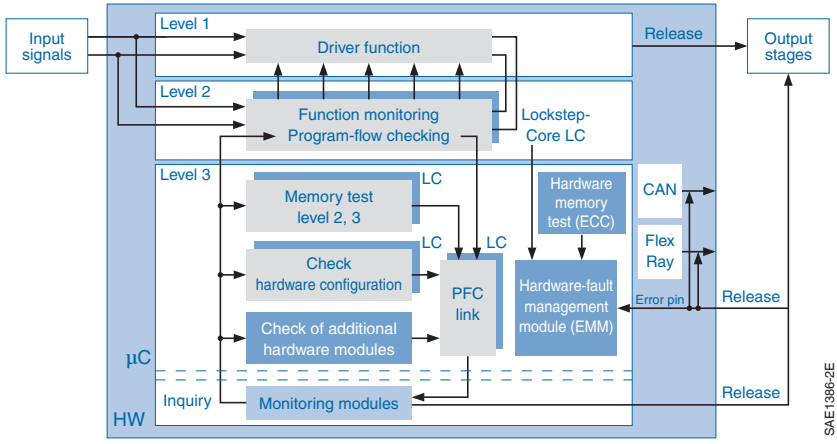
Using the worldwide AUTOSAR standard (Automotive Open System Architecture [5], see Software engineering) ensures a high degree of software reusability and exchangeability. AUTOSAR takes into consideration the complete vehicle system, not just an individual ECU.

Application

ECUs are, as described above, structurally always very similar (Input – Processing – Output). The specific control required for the respective application is performed by the program executed in the microcontroller. This program consists of an executable program code and a data part (individual characteristics, characteristic curves, and program maps), the so-called application data. The entire program is stored in the read-only memory, designed in ECUs as a flash memory. To satisfy the different requirements for controlling for example a driving engine and the statutory provisions for exhaust-gas emissions and diagnostics, it is essential to change and optimize the application data.

Whereas roughly 100 to 1,000 application data were sufficient for the first generations of ECUs for engine management, today several 10,000 are required. The process of optimizing application data is called “application of ECUs”. The work necessary for this is carried out on test benches and in test vehicles and requires specially prepared engine control units.

Figure 12: ECU-software levels and hardware-based monitoring module



Functional safety

As the complexity of electronic components in vehicles constantly increases, so too does the potential for malfunctions. If a safety-relevant component is affected by such a malfunction, people may come to harm in the worst cases [3]. The engine control unit uses hardware and software functions to monitor itself continuously for malfunctions. Three software levels and a hardware-based monitoring module are implemented for this purpose (Figure 12).

The level 1 software contains the desired application functions. These check the plausibility of input signals supplied by different sensors and thereby blank out invalid sensors and implement appropriate substitute reactions. The level 2 software calculates in parallel in an independent computer core parts of the application software in a redundant, independent implementation (e.g., generation of torque or triggering the starter). The output stages are enabled only when both levels arrive at the same result. The level 3 software checks hardware-related faults (e.g., damaged memories).

On the hardware level the ECU has a monitoring module. Microcontroller and monitoring module monitor each other by way of a "question-and-answer game" (for example solving a computing task, the monitoring module specifies a number which must be incremented by the microcontroller). If a fault is detected, both can initiate appropriate substitute functions independently of one another.

The effectiveness of the switch-off paths of levels 2 and 3 and of the monitoring module is checked each time the ECU is switched on.

EOL programming

The large number of vehicle variants with different control programs and data records calls for a procedure for reducing the ECU types required by the vehicle manufacturer. To this end the complete memory area of the flash EPROM can be programmed with the ECU program and the variant-specific data record at the end of vehicle production with EOL programming (End of Line). Data transfer is effected via the communication interface.

Alternatively, to reduce the variant diversity in the memory, it is possible to store several data variants which are then selected by means of coding at the end of line. This coding is stored in the EEPROM.

References

[1] ISO 16750: Road vehicles – Environmental conditions and testing for electrical and electronic equipment.

[2] G.E. Moore: Cramming more components onto integrated circuits. Electronics 38, 114–117 (1965).

[3] ISO 26262 Road vehicles – Functional safety.

[4] H. Randoll: Elektronische Steuerung von Dieselmotoren. In: K. Mollenhauer, H. Tschöke, R. Maier (Editors): Handbuch Dieselmotoren, 4th Ed., pp. 365–371. Verlag Springer Vieweg (2018).

[5] Automotive Open System Architecture (2018), https://www.autosar.org/.

Mechatronics

Mechatronic systems and components

Definition

The term “mechatronics” is a compound derived from the words mechanisms and electronics, where electronics means “hardware” and “software”, and mechanisms is the generic term for the disciplines of “mechanical engineering” and “hydraulics”. It is not a question of replacing mechanical engineering by “electronification”, but of developing a synergistic approach and design methodology (Figure 1). The aim is to achieve a synergistic optimization of mechanical engineering, electronic hardware and software in order to project more functions at low cost, less weight and installation space, and better quality.

A crucial factor governing the success of a mechatronic approach to solving problems is to regard the two previously separate disciplines as a single entity.

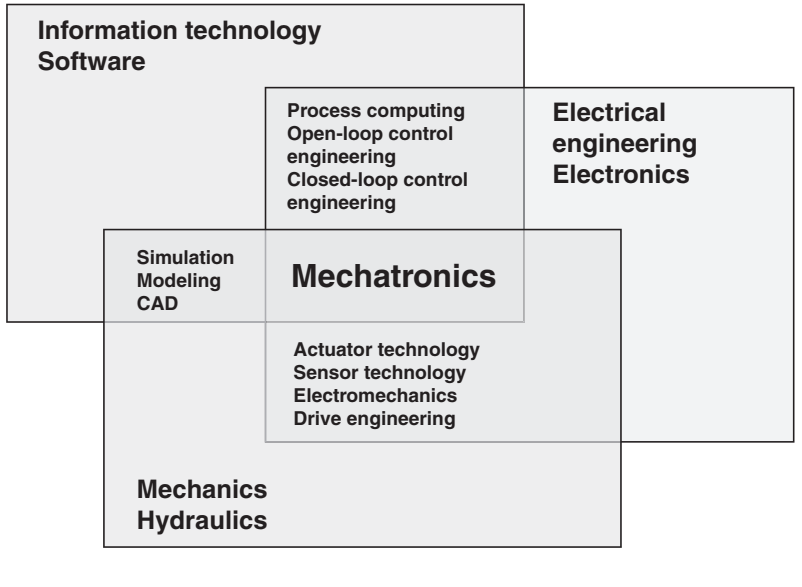
Application

Mechatronic systems and components are now used in practically all aspects of the automobile today, starting with engine management and fuel injection for gasoline and diesel engines, transmission-shift control, and electrical and thermal energy management, through to a wide variety of braking and vehicle-dynamic systems. It even includes communication and information systems, with many different requirements when it comes to operability. Besides systems and components, mechatronics are also playing an increasingly vital role in the field of micromechanics.

Examples at system level

A general trend is emerging in the advanced development of systems for fully automated vehicle handling and guidance: In future, mechanical systems will be replaced to an increasing extent by “x-by-wire” systems. One system that has been

Figure 1: Interactions of mechatronics



in existence for quite some time is "drive-by-wire", i.e. electronic throttle control. "Brake-by-wire" is replacing the mechanical and hydraulic link between the brake pedal and the wheel brake. Sensors detect the driver's brake command and transmit this information to an electronic control unit. The unit then generates the required braking effect at the wheels by means of actuators.

A possible option for implementing "brake-by-wire" is the electrohydraulic braking system (SBC, Sensotronic Brake Control). When the brake is operated or in the event of stabilization interventions by driving-dynamics control, the SBC ECU calculates the desired setpoint brake pressures at the individual wheels. Since the unit calculates the required braking pressures separately for each wheel and detects the actual values separately, it can also regulate the braking pressure to each wheel via the wheel-pressure modulators. The four pressure modulators each consist of an inlet and an outlet valve controlled by electronic output stages which together produce a finely metered pressure regulation.

In the common-rail system, pressure generation and fuel injection are separated from each other. A high-pressure accumulator, i.e. the common rail, stores constantly the fuel pressure required for each of the engine's operating states. A solenoid-valve-controlled injector with integrated nozzle assumes the function of injecting the fuel directly into the combustion chamber of each cylinder. The engine electronics constantly request data on accelerator-pedal position, rotational speed, operating temperature, fresh-air intake flow, and rail pressure in order to optimize the control of fuel metering as a function of the operating conditions.

Examples at component level

Fuel injectors are crucial components in determining future potential of diesel-engine technology. Common-rail injectors are an excellent example of the fact that an extremely high degree of functionality and, ultimately, customer utility can only be achieved by controlling all the physical domains (electrodynamics, mechanical engineering, fluid dynamics) to which these components are subjected.

In-vehicle CD drives are exposed to particularly tough conditions. Besides wide temperature ranges, they must withstand extreme vibrations that have a critical impact on such precision-engineered systems. The drives are usually equipped with a spring-damper system to isolate the playback unit from vibrations that occur when the vehicle is moving. Any considerations to reduce the weight and installation space of CD drives immediately raise questions concerning these spring-damper systems. If the damper system is eliminated from a CD drive, the main focus is on designing a mechanical system with zero clearances and producing additional reinforcement for the focus and tracking controllers at high frequencies. Only by considering both measures from a mechatronic viewpoint is it possible to achieve an optimized vibration-proof solution for an automotive environment. Besides weight savings of about 15 %, the installation height has also been reduced by about 20 %.

The new mechatronic approach for electrically driven coolant motors is based on brushless, electronically commutated DC motors. Initially, they are more expensive (motor with electronics) than previous motors equipped with brushes. However, the overall optimization approach has a positive tradeoff: Brushless DC motors can be used as "wet rotors" with a much simpler design. This reduces the number of single parts by roughly 60 %. Taking the aggregate view, the sturdier design has double the service life, almost half the weight, about 40 % less the overall length, while maintaining costs at a comparable level.

Examples in the field of micromechanics

Another area of applications for mechatronics is the field of micromechanical sensors, with noteworthy examples such as hot-film air-mass meters and yaw-rate sensors.

The design of microsystems also requires an interdisciplinary approach owing to the close interaction between the subsystems involving individual disciplines such as mechanical engineering, electrostatics, fluid dynamics (where necessary), and electronics.

Development methodology

Simulation

The special challenges that designers face when developing mechatronic systems are the ever shorter development times and the increasing complexity of the systems. At the same time it is vital to ensure that the developments will result in useful products.

Complex mechatronic systems consist of a large number of components from a wide range of physical domains: hydraulics, mechanical engineering, and electronics. The interaction between these domains is a decisive factor governing the function and performance of the overall system. Simulation models are required to review key design decisions, especially in the early development stages when there is no prototype available.

Basic issues can often be clarified by producing relatively simple models of the components. If more detail is required, more refined component models are needed. The detailed models focus mainly on a specific physical domain.

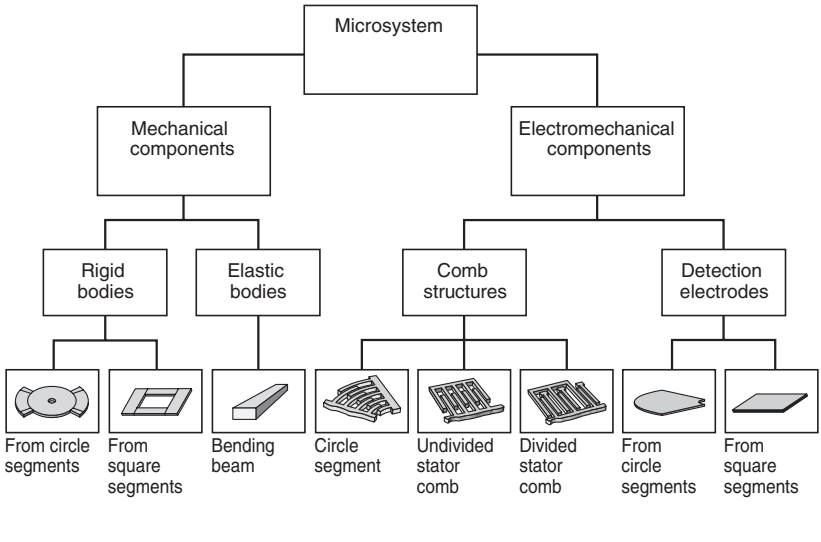
As a result, there are detailed hydraulic models of common-rail injectors. They are simulated using special programs whose algorithms are matched precisely to the hydraulic systems. For example, the requirements here would be to take cavitation phenomena into consideration.

Detailed models are also needed to design the power electronics for activating the injectors. Again, this involves the use of simulation tools which must be developed specifically to design electronic circuits.

Tools that are specially designed for this specific part of the overall system are also required to develop and simulate the control-unit software which controls the high-pressure pump and power electronics using signals from the sensors.

As the components in the overall system interact with each other, it is not sufficient to consider specific detailed models of the components in isolation. The optimum solution is also to take into account the models of other system components. In most cases, these components can be portrayed by much simpler models. For

Figure 2: Model library for a micromechanical yaw-rate sensor



example, a system simulation focused on hydraulics only requires a simple model of the power electronics.

The application of various domain-specific simulation tools during the design of mechatronic systems is only efficient if there is some sort of support for exchanging models and parameters between the simulation tools. The direct exchange of models is highly problematic due to the specific languages used for describing the models of each of the tools.

However, an analysis of the typical components in mechatronic systems shows that they can be composed of a few simple elements specific to the domains. These standard elements include, for example:

– in hydraulics: restrictor, valve, or pipe,
– in electronics: resistor, capacitor, or transistor,
– in mechanical engineering: mass with friction, transmission, or clutch (accordingly for micromechanics).

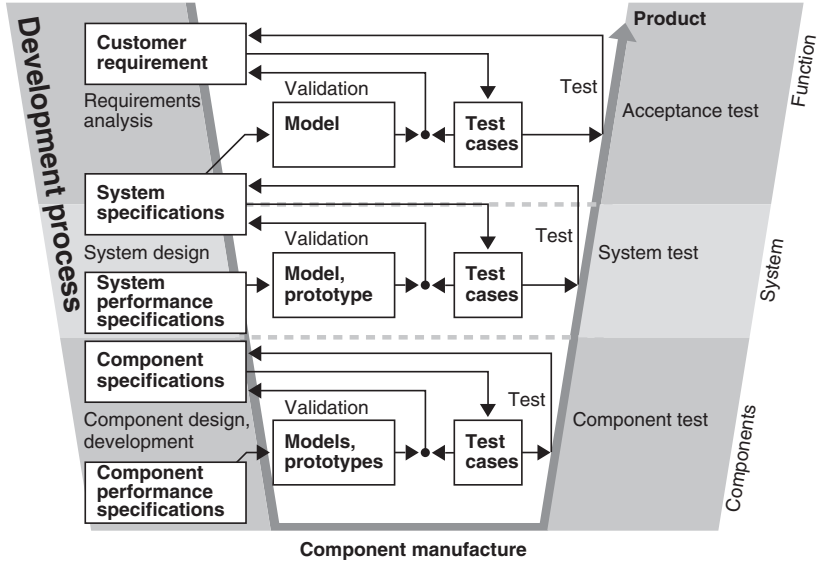
The preferable solution is that these elements should be stored in a central standard model library (Figure 2) that is also decentrally accessible to product development. The kernel of the standard model library is a documentation of all the standard elements. For each element, this comprises:

– a textual description of the physical behavior,
– the physical equations, parameters (e.g. conductivity or permeability), and state variables (e.g. current, voltage, magnetic flux, pressure),
– a description of the associated interfaces.

V model

The relationships between the different phases of product development are illustrated in the V model (Figure 3): from requirements analysis through drafting, development, implementation, and testing through to system introduction of the

Figure 3: General overview of V model

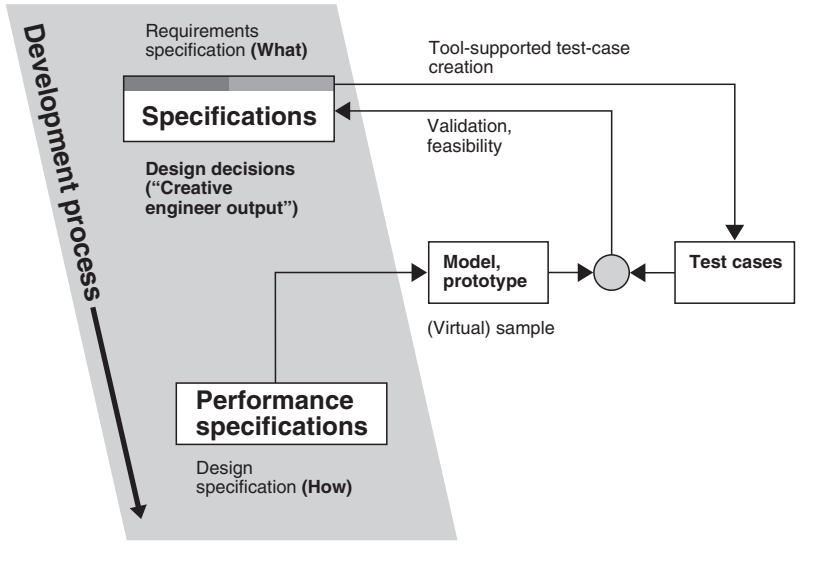


product. A project passes through three "top-down" levels during the development stage:
– customer-specific functions,
– system, and
– components.

A requirements specification ("what") must first be produced at each level in the form of specifications (Figure 4). This is then used to produce the design specifications based on design decisions (the actual creative engineering output). The performance specifications describe "how" a requirement can be met. The performance specs form the basis for a model description which allows a review (i.e. validation) of the correctness of each design stage together with previously defined test cases. This procedure passes through each of three stages, and, depending on the technologies applied, for each of the associated domains (mechanical engineering, hydraulics, fluid dynamics, electrics and electronics, software).

Recursions at each of the design levels shorten the development stages significantly. Simulations, rapid prototyping, and simultaneous engineering are tools that allow rapid verification, and they create the conditions for shortening product cycles.

Figure 4: Recursion methodology at one level



Outlook

The driving force behind mechatronics is the continuous progress in microelectronics. Mechatronics benefits from computer technology in the form of ever more powerful integrated computers in standard applications. Accordingly, there is a huge potential for further increases in safety and comfort in motor vehicles, accompanied by further reductions in pollutant emissions and fuel consumption. On the other hand, engineers face new challenges in mastering the new technologies for these systems.

Even in the event of a fault, future "X-by-wire" systems must continue to be capable of fulfilling a prescribed functionality without reverting to a mechanical or hydraulic fallback level. The condition for their implementation is a high-reliability and high-availability mechatronic architecture which requires a "simple" proof of safety. This affects both single components as well as energy and signal transmissions.

Besides "x-by-wire" systems, driver-assistance systems and their related man-machine interfaces are another field in which significant progress can be achieved for users and automotive manufacturers by the systematic implementation of mechatronic systems.

The design approaches of mechatronic systems should strive toward continuity in several aspects:

- Vertical: "Top-down" from system simulation, with the objective of overall optimization, through to finite-element simulation to achieve a detailed understanding, and "bottom-up" design engineering from component testing through to system testing.
- Horizontal: "Simultaneous engineering" across several disciplines in order to deal with all product-related aspects at the same time.
- Across corporate boundaries: The concept of a "virtual sample" is approaching gradually step by step.

Another challenge is training in order to further an interdisciplinary mindset and develop suitable DE processes and forms of organization and communication.

Automotive software engineering

Motivation

Aim of development
Any development has the aim of creating a new function or improving an existing function of the vehicle. Such functions bring about an added value for the user of the vehicle (e.g. occupants, mechanics in the workshop or haulage contractors), compliance with statutory requirements, simplified maintenance, or improved development or manufacturing efficiency. Technical implementation can be mechanical, hydraulic, electrical or electronic. Often it is combinations of these technologies, where electronics increasingly plays a key role in the realization of many innovations in a motor vehicle. Thanks to the use of electrics, electronics and software – the logical core of systems – "intelligent" functions of the drive, the chassis and the rest of the vehicle are cost-effectively realized.

Virtually all vehicle functions in all vehicle classes are now electronically controlled or monitored. Ongoing advances in the technology and performance of electronics hardware allow the realization of many new, more powerful functions through the use of software. The various electronic systems are being increasingly cross-linked in the vehicle. Efforts are pursuing cross-linking vehicles to each other and with the environment (e.g. Internet).

Software requirements
The specific requirements of the software are highly varied. Many systems for the engine and for driving safety must operate with "real-time capability", i.e. the response of the control operation must keep pace with the physical process. When very fast physical processes are being controlled, such as for example engine management or driving-dynamics control, the calculation must therefore be made very quickly. Even the requirements with regard to reliability are high in many fields. This is particularly true of safety-relevant functions. The software and the electronics are monitored by complex diagnostics.

The software is used across the many variants of a vehicle or even across model series. It must then be adaptable to the corresponding destination system. It contains calibration parameters and program maps for this purpose. These amount to several 10,000 per vehicle. These manipulated variables are dependent on each other in many ways. In addition, a functionality is distributed beyond multiple systems or control units more and more frequently.

The software is for the most part developed specially for the corresponding applications and is integrated into the entire system. It is called "embedded software". The many functions are often developed and further developed over a long period of time in many locations around the world. Because spare parts must remain available even after the production of a particular vehicle has come to an end, the electronics in the vehicle must have a relatively long life cycle of up to 30 years.

For cost reasons, ECUs often contain microcontrollers with limited computing power and limited memory space. This requires optimization measures in software development to reduce the hardware resources required.

The characteristic features of software differ depending on the field of application. While the software for the drivetrain is extensive, the emphasis is on real-time performance in the chassis applications. In safety and comfort and convenience applications, the emphasis is on efficiency, i.e. the consumption of resources, and in multimedia applications large volumes of data must be processed in a short period of time.

The complexity resulting from these requirements and features must be economically mastered in a cooperation of development between vehicle manufacturers and suppliers.

Design of software in motor vehicles

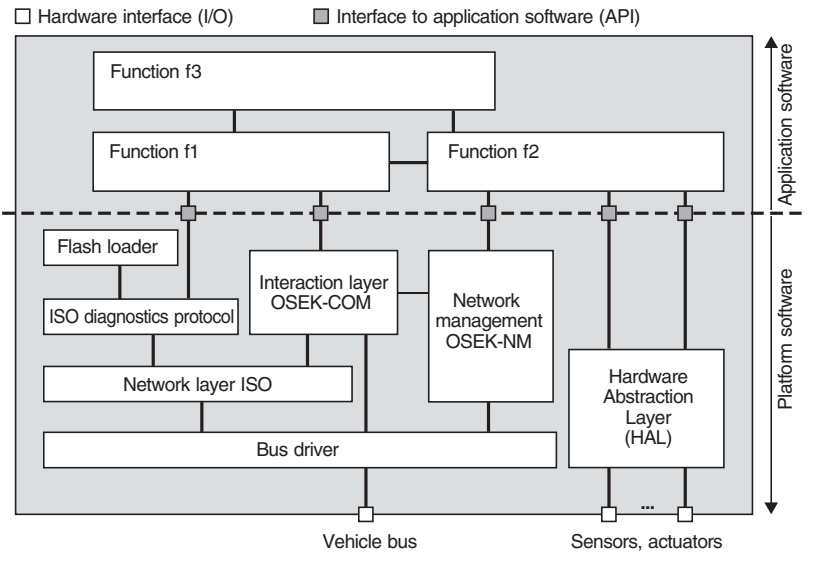
The software in a motor vehicle is made up of many components. A distinction is essentially made, as is the case with a PC, between the "perceptible functions" of the software, the application software and platform software partly dependent on the hardware (Figure 1). The interaction between all the functions is defined in the architecture. Different views can be adopted here. The static view describes hierarchically the function groups, the signals and the distribution of the resources. On the other hand, the functional view describes the signal progression through the different functions. The dynamic, i.e. time-dependent view, considers the time response in the execution of the various tasks. Standards have already been introduced at an early stage to ensure the interaction between and further development of the individual components. The most important standards are explained in the following.

Important standards for software in motor vehicles

Bodies/committees

The "Association for Standardization of Automation and Measuring Systems" (ASAM) is a standardization body in the automotive industry for data models, interfaces and syntax specifications [5]. The ASAM developed different standards for connecting an ECU to a computer or a data-entry terminal. The standard ASAM-MCD1 (MCD stands for Measurement, Calibration and Diagnosis) supports different data transport protocols for this purpose. By using ASAM-MCD2 specifications, the binary data in the ECU can be addressed and the relevant quantities can at the same time be displayed in connected tools as physical values and processed. The standard ASAM-MCD3 also permits an automation of such processes, for example for the automatic calibration of data records. Further ASAM standards cover, for example, the exchange of functional descriptions and data.

Figure 1: Main component parts of the static software architecture for microcontrollers and standardized software components



The FlexRay Consortium developed the specification for the FlexRay field bus for open and closed-loop control in the automobile field. Thanks to high transfer rates with predefined bus arbitration and a fault-tolerant design it is particularly suitable for use in active safety systems and in the drivetrain (see FlexRay).

The International Electrotechnical Commission (IEC) is an international standardization body in the fields of electrical engineering and electronics [6]. The IEC offers three evaluation systems with which conformity to the international standards can be verified. The IEC works in close cooperation with the International Organization for Standardization (ISO) [7], the International Telecommunication Union (ITU) and numerous standardization bodies (including the Institute of Electrical and Electronics Engineers, IEEE) [8].

The Motor Industry Software Reliability Association (MISRA) is an organization in the automotive industry that devises rules for the reliable development and application of software in vehicle systems [9]. The most widely known is the programming standard MISRA-C, which was developed by the MISRA. It prescribes programming rules for reliable programming in the programming language C. The purpose of this standard is to avoid runtime errors due to unreliable C constructs and structural weaknesses due to misunderstandings between programmers, and to safeguard the validity of expressions. Many rules can be automatically checked and taken into account in the generation of codes.

The Society of Automotive Engineers (SAE) is an international science and technology organization operating in the field of mobility technology [10]. It sets among other things standards for the automotive industry and promotes the exchange of knowledge and ideas.

The standardization body "Open Systems and their Interfaces for the Electronics in Motor Vehicles" (Offene Systeme und deren Schnittstellen für die Elektronik im Kraftfahrzeug, or OSEK) came about from a project by the German automotive industry. Later came the "Vehicle Distributed Executive" (VDX) initiative of the French automotive industry. Standardizations of basic-software components are established under the term OSEK/VDX in the following fields [11]:

- Communication (data exchange within and between ECUs)
- Operating system (real-time execution of ECU software and basic services for other OSEK/VDX modules)
- Network management (configuration and monitoring).

The Japan Automotive Software Platform and Architecture (JasPar) is an initiative for reducing costs and developing technology in automotive electronics. It encourages Japanese companies to develop jointly non-competition-relevant technologies such as network solutions, service functions and basic software. JasPar works in close cooperation with AUTOSAR and FlexRay.

AUTOSAR

AUTOSAR (Automotive Open System Architecture) [12] is a development partnership of vehicle manufacturers, ECU manufacturers and manufacturers of development tools, ECU basic software and microcontrollers. The aim of AUTOSAR is to simplify the exchange of software on different ECUs. To this end, a standardized software architecture has been devised with standardized description and configuration formats for embedded software in automobiles. AUTOSAR defines methods for describing software in vehicles which ensure that software components can be reused, exchanged, scaled and integrated. AUTOSAR is becoming accepted by an increasing number of automakers.

Crucial to AUTOSAR is a logical distribution into ECU-specific basic software (BSW) and ECU-independent application software (ASW) and their connection via the virtual function bus system (VFB) (Figure 2). This virtual function bus also connects software components implemented in different ECUs. In this way, these can be shifted between different ECUs without changes having to be made in the affected software components themselves. This can be useful in optimizing computing power, memory requirements or communication load.

The functional software components (SWC) are strictly separated from each other and from the basic software. They typically contain specific control algorithms executed at run time, the “runnable entities”. They communicate via the AUTOSAR interface with other functions and the ECU Interfaces. These interfaces (API) are defined in SWC XML descriptions.

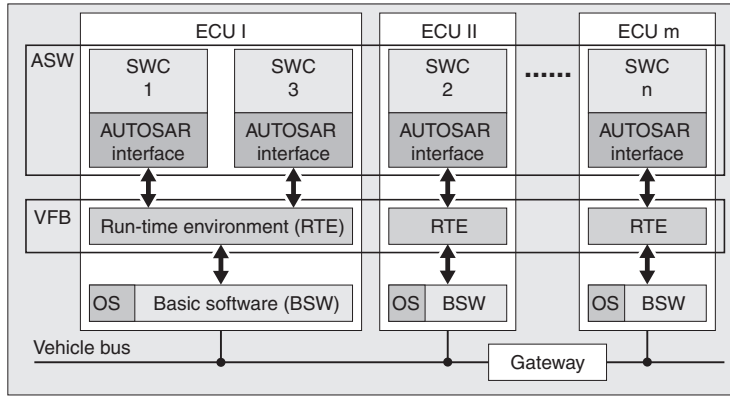
The run-time environment (RTE) provides the communication services between the functional software components and the corresponding basic software on the ECU. The RTE is tailored to the specific ECU and the application. It can be generated to a large extent automatically from the interface requirements.

The basic software contains the ECU-specific program parts, such as the communication interfaces, diagnostics and memory management. The basic software also contains the service layer. This software combines software components for general service functions (SRV), communication (COM) and the operating system partly dependent on the ECU used (OS) [4, 11]. The latter is based on the OSEK/VDX OS. In this field, the resources of the ECU are subdivided and managed so as to arrive at optimal network support, memory management, diagnostics, etc.

The hardware used is encapsulated in two layers based on each other. Abstraction of the microprocessor (Micro-Controller Abstraction Layer, MCAL) with direct access to the interface modules of the ECU continues in a further layer (ECU Abstraction). “Complex Device Drivers” (CDD) facilitate direct access to microcontroller resources for applications with special functionality and timing requirements. They are also an integral part of the basic software such that the application software can be developed independently of the hardware, even when the services of the complex device drivers are required.

In addition to the ECU architecture, the development methods are also standardized in part by AUTOSAR. This relates, above all, to the structure and the dependencies of the different work products (e.g. files). These are needed in order to generate from the different software component descriptions executable programs for the respective ECUs.

Figure 2: AUTOSAR architecture
ECU Electronic control unit, ASW Application software, SWC Software components, VFB Virtual function bus, RTE Run-time environment, BSW Basic software, OS Operating system.



Diagnostic standards
Vehicle-specific systems for diagnostics for vehicle development and production and for workshops prove to be high-maintenance, cost-intensive and inflexible. They tie the manufacturers to suppliers and prevent simple data communication in the case of cross-company cooperations. Some diagnostic standards have therefore been developed (e.g. [13] and [14]).

The Automotive Electronics working group (ASAM-AE) [5] has devised three specifications for data-based (i.e. software-based) vehicle diagnostics which are published as international standards in the ISO 22900 standard group [15]:
– Interface between run-time environment and communication hardware (MCD-1D and PDU-API, ISO 22900).
– ODX standard for exchange of diagnostic data, e.g. for supplying the workshop tester with data (MCD-2D, ISO 22901, [16]).
– Object-oriented programming interface (MCD-3D, ISO 22900) for diagnostic applications, such as e.g. prompted troubleshooting.

The MCD-1D standard takes into consideration existing standard tools such as, for example, devices for flashing ECUs.

Requirements with regard to an exchange format, the "Open Test Sequence Exchange Format" (OTX), are currently being defined within the framework of an ISO draft in order to create, use and exchange diagnostic sequences.

The development process

The focal point of software development is the depiction of the logical system architecture in a concrete software system with all programs and data. Here, the entire processor-controlled system of the vehicle is considered. Particular importance is attached to a clear separation of specification, design and implementation. Specification of the software functions is performed on the physical level, while design and implementation of programs and data are geared towards the specific microcontroller.

In order to satisfy the above-mentioned requirements in the development of software in automobiles, the defined sequences (processes) are, in addition to the technology and the tools, an integral part of development.

Process description models
Numerous more or less complicated models are used to describe the workflows in software development. They serve to make the sequences transparent, to compare them, to identify problem areas, and to verify conformity in accordance with defined standards. They were, however, not originally conceived to improve directly the quality of the software itself, to increase efficiency or to eliminate systematic faults in the sequences. Process description models are therefore only partly suitable. The common V model will be described here by way of example.

Principle of the V model
The V-shaped representation of the development sequence described here is used in many variations and degrees of detail. The "V model" of the federation for planning and implementing IT projects of central government in Germany [17] will not be described here.

The V model divides the process steps associated directly with development along a V, where the x-axis depicts the development progress and the y-axis the depth, i.e. the degree of detail, of the corresponding process step (Figure 3). A process step can be described by the required input variables, the procedure, the methods, the roles, the tools, the quality criteria, and the output variables. The process steps defined on the left arm are verified on the right arm. These steps can also run several times or be divided.

In the extended V model, accompanying processes such as, for example, request, change, project and quality management, can be considered.

Process evaluation models

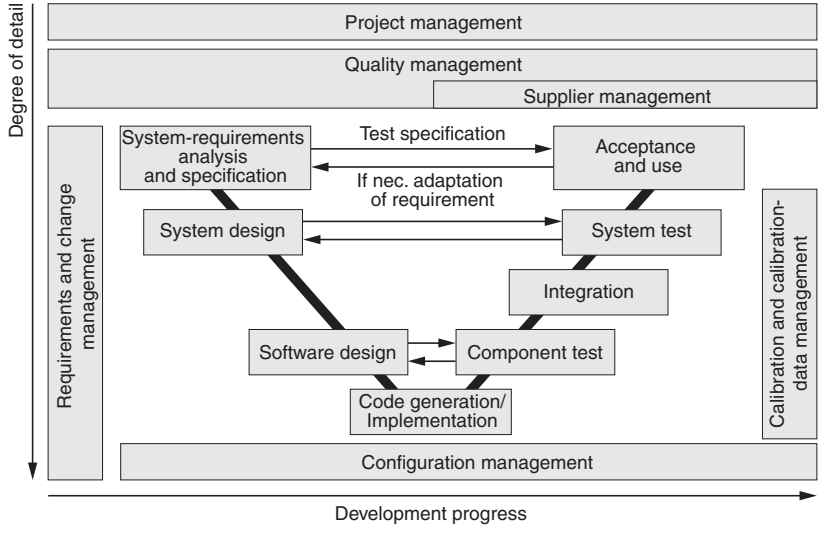
Process evaluation models provide, in addition to a pure description of the tasks and sequences, information on the maturity and the quality of the processes. Work steps can thus be compared, evaluated and certified. In this way, it is also possible to identify process gaps which impact, for example, on the product quality or the costs. However, the information on the quality of the processes does not provide a complete picture of the quality of the products themselves here either. The three most important process evaluation models will be described here.

ISO 9000 and ISO/TS 16949

The process-oriented EN ISO 9000 standard series [18] and following stipulates the requirements with regard to a quality management system. The focus here is on the interactions and the interfaces. The original emphasis was on production and the customer interfaces.

The ISO/TS 16949 technical specification [19] was devised in the North American and European automotive industries and standardizes requirements relating to quality management systems. The aim of the standard is to improve effectively system and process quality in order to increase customer satisfaction, to identify failures and risks in the production process and in the supply chain, to eliminate their causes, and to check the effectiveness of the corrective and preventive ac-

Figure 3: Extended V model



tions taken. The core of the specification is not the discovery, but rather the avoidance of failures.

Compliance with ISO 9000 and ISO/TS 16949 can be verified by certification.

CMMI

"Capability Maturity Model Integration" (CMMI) is a model for evaluating and systematically improving development organizations and their processes [20] which was originally developed by the Software Engineering Institute (SEI). It describes a collection of requirements with regard to the processes and their dependencies (Figure 4). CMMI offers a framework, the implementation of which calls for a business-oriented interpretation and organization of content. It describes what has to be done. The organization must shape the "how" appropriately. The content of CMMI is based on essential industry "best practices". CMMI provides a procedure for process improvement created in the longer term from the path of organization development through to learning organization.

CMMI has much in common in terms of content with ISO 9000 and ISO TS 16949. In this context, CMMI has a higher degree of detail, while ISO 9000 and ISO TS 16949 cover a wider range of application.

CMMI distinguishes between five maturity levels (ML) for an organization unit (Figs. 4, 5). Different process areas are considered, depending on the maturity

Figure 4: CMMI process overview
ML Maturity level

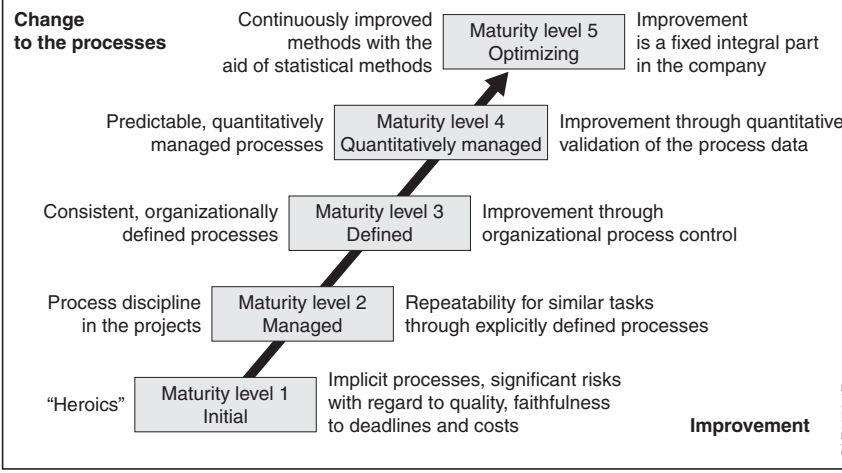
Category	Process areas
Process management	ML3 Organizational process focus (OPF) ML3 Organizational process definition (OPD) ML3 Organizational training (OT) ML4 Organizational process performance (OPP) ML5 Organizational innovation & deployment (OID)
Project management	ML2 Project planning (PP) ML2 Project monitoring and control (PMC) ML2 Supplier agreement management (SAM) ML3 Integrated project management (IPM) ML3 Risk management (RSKM) ML3 Quantitative project management (IPM)
Engineering	ML2 Requirements management (REQM) ML3 Requirements development (RD) ML3 Technical solution (TS) ML3 Product integration (PI) ML3 Verification (VER) ML3 Validation (VAL)
Support processes	ML2 Configuration management (CM) ML2 Process and product quality assurance (PPQA) ML2 Measurement and analysis (MA) ML5 Cause analysis and rectification (CAR) ML3 Decision analysis and resolution (DAR)

Process areas

Elementary management and support processes (ML2)	Engineering processes (ML3)
Organization processes and further management processes (ML3)	"High-maturity" processes (ML4 and ML5)

SAE1177-2E

Figure 5: CMMI maturity level



level. A maturity level is reached when all the corresponding process areas are mastered and verified by an assessment. To reach a higher maturity level, the process areas of the maturity levels underneath must also be verified again.

CMMI is used as an improvement and evaluation model for development organizations and offers good support for organization-wide process optimization and auditing of suppliers.

Automotive SPICE

SPICE stands for "Software Process Improvement and Capability Determination". Automotive SPICE [21] is a motor-vehicle-specific variant of the international ISO/IEC 15504 standard (processes in the life cycle of software, [22]). It is a model for project-specific evaluation of software development processes and concentrates, like CMMI, on the requirements for systematic development. The model content of Automotive SPICE and of CMMI is therefore very similar. Automotive SPICE concentrates more on the requirements level and affords less scope for the business-oriented interpretation of requirements. Automotive SPICE focuses (currently) only on the software and only on individual projects. In contrast, the ranges of application of CMMI go further, comprising development activities and services of all kinds and their direction through the organization. Automotive SPICE is used by motor-vehicle manufacturers as an evaluation model for software projects of suppliers.

The process assessments are carried out using the two-dimensional reference and assessment model. The "process dimension" serves to identify and select the processes to be analyzed in the assessment, the "maturity-level dimension" on the other hand serves to determine and evaluate their respective capability. The maturity-level dimension consists of the six maturity-level stages "incomplete", "performed", "controlled", "established", "predictable" and "optimizing".

Quality assurance in software development

As with any technical product, numerous tools for quality assurance are also used on software. In contrast to mechanics and electrics, quality assurance plays a lesser role in the manufacture of software, since software is relatively easy to reproduce. Important emphases are the overall functionality of the system, the quality standards, mastery of complexity, and application. Because the software in a motor vehicle also includes safety-relevant systems such as, for example, driving-dynamics and driver-assistance systems, the verifiability of quality plays an important role. Even the economic representation of the aimed-for software quality, especially in complex systems, is very important.

ISO 26262

Based on the IEC 61508 standard [23], the ISO 26262 standard [24] has been introduced by the automotive industry for drafting safety-related electric and electronic systems in motor vehicles. This comprises requirements relating to both the product and the development process, and thus comprises conception, planning, development, implementation, startup, maintenance, modification, shutdown and uninstallation both of the safety-relevant system itself and of the safety-related (risk-reducing) systems. The standard designates these phases in their entirety as the "entire safety life cycle". The products are subdivided into Safety Integrity Level SIL 1 through SIL 4 (ISO 61508) and Automotive SIL, ASIL A through ASIL D (ISO 26262). SIL 1 and ASIL A are the lowest, SIL 4 and ASIL D the highest safety integrity level.

Workflows of software development in motor vehicles

Interdisciplinary cooperation in development (e.g. between drive and electronics development), distributed development (e.g. between supplier and vehicle manufacturer or at different development locations) and the long life cycles of the software elements call for a common and holistic understanding of the tasks involved. For example, when designing the open-loop and closed-loop control functions of a vehicle, it is also necessary to maintain an overview of overall reliability and safety requirements, as well as the aspects of software realization. To meet these challenges, model-based development has established itself in many fields of application (Figure 6).

Model-based development

Model-based development distinguishes between two areas. The logical system architecture comprises and describes the virtual area of the models, while the technical system architecture contains the real ECUs and vehicles. The "logical system architecture" is gray in the figures and the "technical system architecture" is white. This procedure is described in terms of open-loop and closed-loop control functions, but is also suitable for general realization of functions – for example, for monitoring and diagnostic functions.

A graphics-based function model considers all system components and can be used as the basis for a general understanding. In software development, the use of model-based development methods, with notations such as block diagrams or finite state machines, is replacing written software specifications to an increasing extent. This method of modeling software functions has still other advantages. If the specification model is formally described, i.e. unambiguously and without leeway for interpretation, as a mathematical function, the specification can be executed on a computer in a simulation, and tested quickly and realistically in the vehicle itself by means of "rapid control prototyping". Inconsistencies can also be uncovered more easily.

Methods of automated code generation can be used to implement the specified function models as software components for ECUs. The function models, therefore, need to contain additional software design information. These may include optimization measures, depending on the product characteristics required by the electronic system. Automated code generation also ensures that the code has consistent quality characteristics.

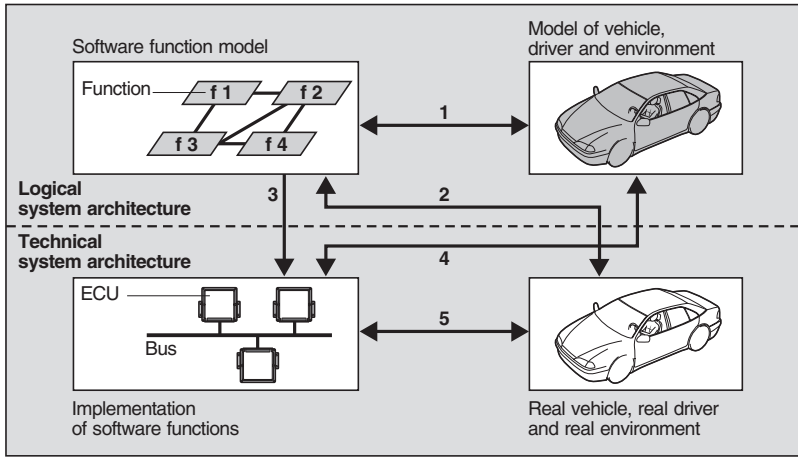
In the next step, virtual environment models, which are supplemented if necessary by real components such as injectors, simulate the environment of the ECUs, facilitating "In-the-loop tests" in the laboratory. Compared to test-bench and road tests, this increases flexibility and enhances test depth, thereby making test cases easier to reproduce.

The calibration of the software functions in the electronic system must consider the vehicle-specific settings, for example, parameters stored in the form of characteristic values, characteristic curves, and characteristic maps of these functions. In many cases, this matching does not take place until a later stage in the development process, frequently directly in the vehicle while the systems are running. The trend, however, is increasingly towards an earlier preliminary data feed; i.e. starting in the early development phases the most realistic calibration data possible is determined using models or empirical values. Because of the multitude of calibration variables and the mutual dependence, the calibration requires suitable procedures and tools, because in the end the quality of the calibration engineering, i.e. the exact adaptation of the software to the vehicle, decides the extent to which the potential of the software is utilized.

Efforts to shift development steps ahead are increasing, i.e. in the realm of simulated (virtual) environment, to detect faults early and save on expensive test hardware and experimental vehicles

Figure 6: Development steps in the model-based development of software functions

Step 1: Modeling and simulation of the software functions for the ECU and of the vehicle, driver and the environment on the computer (software functions are all the functional program parts of the software),
Step 2: Rapid control prototyping of the software functions in the real vehicle,
Step 3: Implementation of the software functions in the real vehicle,
Step 4: Integration and testing of the ECUs with In-the-Loop test systems, laboratory assemblies and test benches,
Step 5: Testing and calibration of the software functions and of the ECUs in the vehicle.



(front loading). High-performance simulation tools and virtual ECU environments make this possible.

Function and ECU networks
This procedure can also be applied to the development of function networks and ECU networks. In these cases, however, additional aspects are introduced, such as:
– combinations of modeled, virtual, and functions already realized in ECU code, and
– combinations of modeled, virtual, and realized mechanical components and hardware.

It is useful, therefore, to distinguish between functions at an abstract level, and technical realization at a more concrete level. The concept of a separate abstract and concrete approach can be applied to all vehicle components, the driver, and the environment.

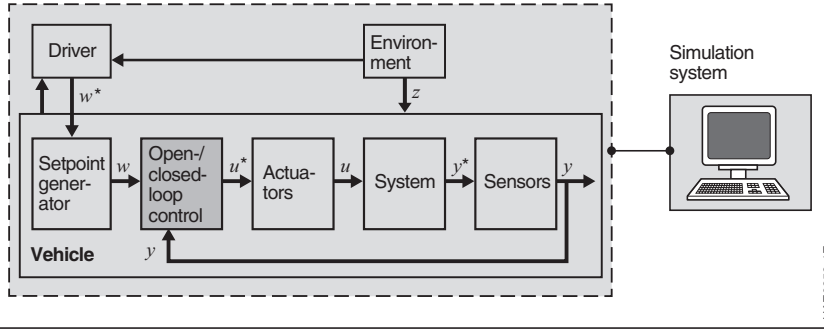
Modeling and simulation of software functions

Modeling
Block diagrams should be used where possible to model open-loop and closed-loop control systems. These diagrams use blocks to represent the response of components, and arrows represent the flow of signals between the blocks (Figure 7). As most of the systems are multivariable systems, all signals are generally in vector form. These are classified into:
– measurement or feedback variables y,
– open- or closed-loop control output variables u^*,
– reference or setpoint variables w,
– driver setpoint values w^*,
– open- or closed-loop control variables y^*,
– manipulated variables u,
– disturbance values z.

The blocks are classified into:
– open-or closed-loop control,
– models of actuators,
– system model,
– models of setpoint generators and sensors, and
– model of the driver and the environment.

The driver can influence the functions of the open- or closed-loop controller by defining setpoint values. All components for entering setpoint values for the driver (for example, switches or pedals), are known as setpoint generators. Sensors, in contrast,

Figure 7: Modeling with block diagrams and simulation



record signals from the plant. A model of this type can be executed on a simulation system (for example, a PC), thus allowing it to be analyzed in more detail.

Rapid control prototyping of software functions

Rapid control prototyping in this context includes all methods for the early implementation of specifications of open-loop and closed-loop control functions in the real vehicle. To allow this, the modeled open- or closed-loop control functions must be implemented in the test. Experimental systems can be used as an implementation platform for the software parts of the open- and closed-loop control functions (Figure 8).

The experimental systems are connected to the setpoint generators, sensors and actuators, as well as the other vehicle ECUs that belong to the overall system. The interfaces to the real vehicle mean that the software functions implemented in the experimental system – and in the ECU – take into account real-time requirements.

Real-time computer systems with considerably higher computing power are usually deployed as ECUs in experimental systems. PCs are increasingly being used as the processor core for this task. This allows the model of a software function to be automatically converted from a specification into an implementable model using a rapid control prototyping tool that is subject to standardized rules. The specified behavior can then be modeled as accurately as possible.

Experimental systems with a modular structure can be configured specifically for the application, for example, in terms of the required interfaces for input and output signals. The overall system is designed for deployment in the vehicle, and is operated using a computer, e.g. a PC. This allows the testing of software function specifications directly in the vehicle at an early stage. The specifications can then be changed as required.

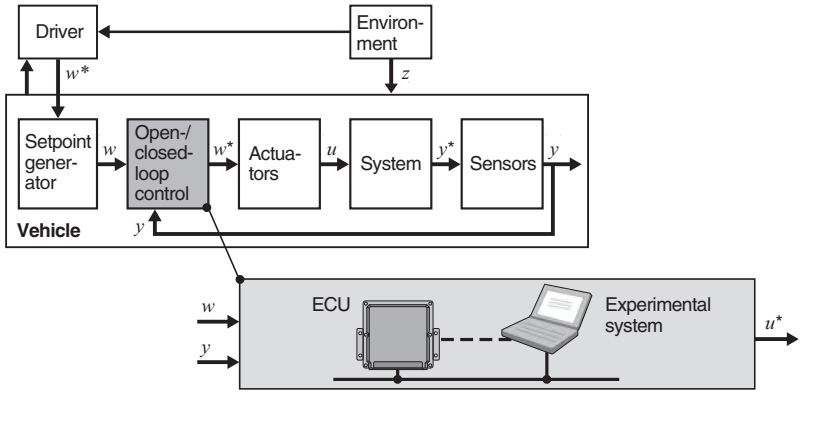
When using experimental systems, there is a choice between bypass and fullpass applications.

Bypass or fullpass applications

Bypass applications are mainly used if only a few software functions are under development, and an ECU with tried-and-true basic functionality is available – for example, from a previous project.

Bypass applications are also suitable if the sensor and actuator functions of an ECU are very complex, and supporting them requires significant effort on the part of the experimental system (for example, in the case of engine ECUs).

Figure 8: Rapid control prototyping for software functions in the real vehicle



Fullpass applications are more suitable if an ECU of this type is not available, if additional setpoint generators, sensors, and actuators also need to be tested, and if the scope of the ECU is of manageable complexity.

It is also possible to combine a bypass of individual software parts and a fullpass of the entire software. This has the advantage of increased flexibility.

Bypass applications

Bypass development is suitable for early testing of an additional or modified software function of an ECU in the vehicle. The new or modified software function is defined by a model and run on the experimental system. This requires an ECU that can run the basic functionality of the software system, support all the necessary desired-value generators, sensors, and actuators, and provide a bypass interface to the experimental system. The new or modified software function is developed using a rapid control prototyping tool. It is then run on the experimental system (Figure 9).

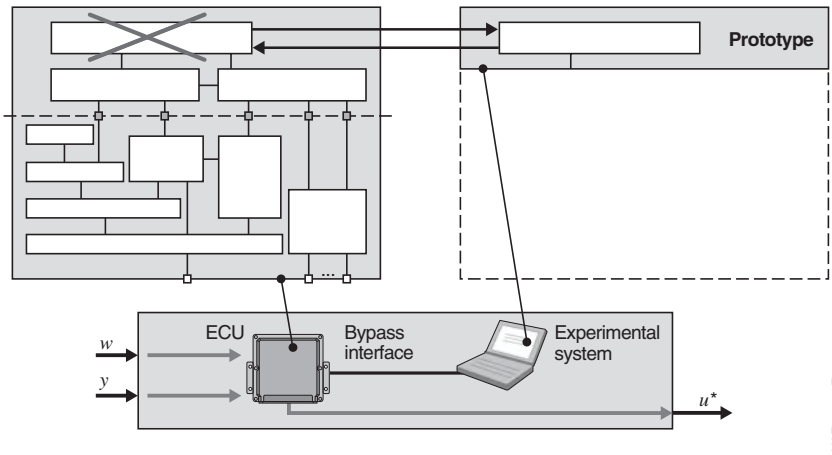
This approach is also suitable for further developments to existing ECU functions. In this case, the existing functions in the ECU are still used, but they are modified to the extent that the input values are sent via the bypass interface, and the output values from the newly developed bypass function are used. The required software modifications to the ECU are called the bypass hooks. These can also be used in previously compiled software with modern development tools. For the essential synchronization of functional computing between the ECU and the experimental system, a procedure is normally adopted in which the ECU triggers computation of the bypass function on the experimental system via a control flow interface. The ECU monitors the output values of the bypass function for plausibility.

The bypass can also be effected via the vehicle bus (e.g. CAN). Even direct access to the CPU of the ECU via the microcontroller interfaces is possible by means of an emulator probe.

Fullpass calibration

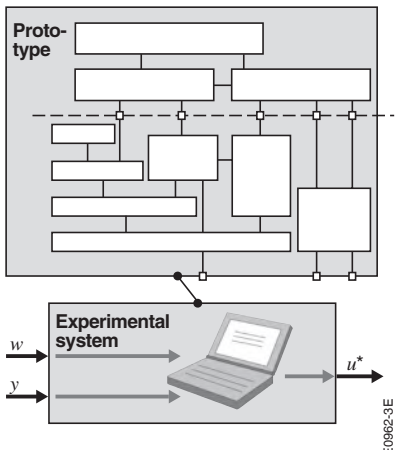
If a completely new function is to be tested in the vehicle, and an ECU with a bypass interface is not available, the test can be carried out using a “fullpass” development. In this case, the experimental system must support all the setpoint generators, sensors, and actuator interfaces for the function. The real-time behavior of the function must also be defined and guaranteed by the experimental system (Figure 10). In general, this is performed by a real-time operating system on the fullpass computer.

Figure 9: Prototype development with bypass system



Virtual prototyping
In the case of complex systems, it is advantageous to test functions as early as possible. One possibility is offered by virtual prototyping. Here the prototype is tested on a virtual environment model. The operating system of the later ECU (e.g. RTA) runs on the experimental system. This enables the time response of the later software to be considered as well.

Figure 10: Prototype development with fullpass system



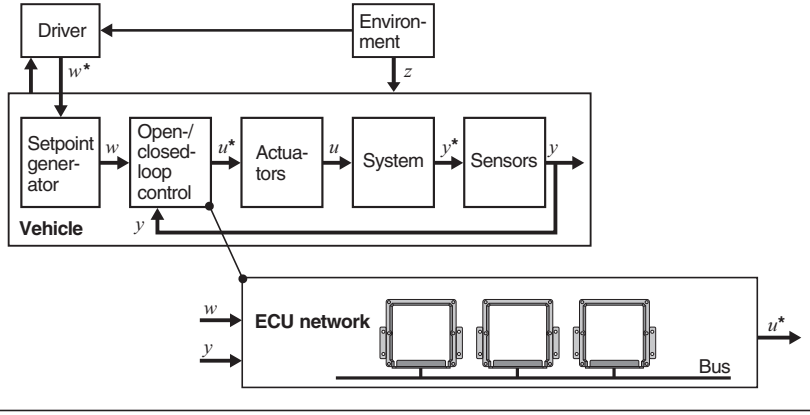
Design and implementation of software functions

Based on the specification of data, the functional behavior and the real-time behavior of a software function, all the technical details of the ECU network, the implemented microcontroller, and the software architecture must be taken into account at the design stage. The final implementation of the software functions can then be defined and executed on the basis of software components (Figure 11).

In addition to decisions regarding the design and behavior of a software function, that takes the time and discrete-related functions of the microcontrollers into consideration, an implementation includes design decisions regarding real-time behavior, distribution and integration of microcontrollers and ECUs, and reliability and safety requirements of the electronic systems. All requirements for electronic systems and vehicles from the production and service viewpoints must also be taken into account (for example, monitoring and diagnostics concepts, the parameterization of software functions, or software updates for ECUs in the field).

Generation of the code and of the accompanying data (e.g. data for documentation, variant management or preliminary data feed of the calibration) is often performed automatically in accordance with established standards.

Figure 11: Realization of open-loop and closed-loop control functions using an ECU network



Integration and testing of software and ECUs

Requirements

Prototype vehicles are often only available in limited numbers. It means that a component supplier often does not have a complete or up-to-date integration and test environment for the components supplied. Restrictions in the test environment can sometimes limit the possible test steps. For this reason, the environment models often serve as the basis for test systems and test benches in the integration and test phase.

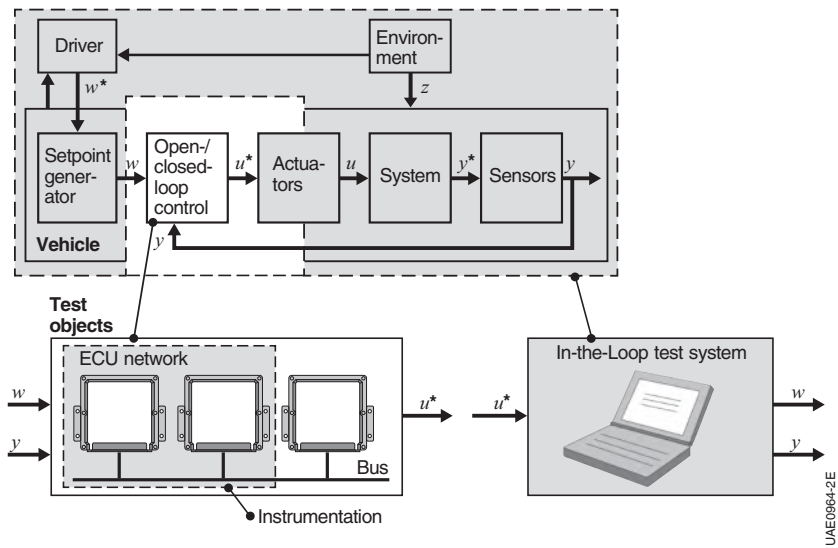
Component integration is a synchronization point for all the individual component developments involved. The integration test, system test, and acceptance test cannot be carried out until all the components are available. For ECUs, this means that the software functions can only be tested when all the components in a vehicle system are available (ECUs, setpoint generators, sensors, actuators, and system). The use of In-the-Loop test systems in the laboratory allows early verifications on ECUs in a virtual test environment in the absence of actual peripheral components (Figure 12).

This allows tests to be performed and automated under reproducible laboratory conditions with a high level of flexibility. In contrast to tests on the test bench or in the real vehicle, the full, unrestricted range of operating states can be tested (for example, an engine ECU can be tested at all load and rpm situations). Vehicle wear and failure situations are easy to simulate, and allow testing of the monitoring, diagnostic, and safety functions of the ECU. Component tolerances (for example, in setpoint generators, sensors, and actuators) can be simulated to allow a verification of the robustness of open-loop and closed-loop control functions.

A significant benefit is also the great test depth which can be achieved through automation. This enables, for example, as many fault types as possible and their combinations to be tested under exactly reproducible conditions and documented by means of a failure log.

This procedure can also be applied to testing actual setpoint generators, sensors, and actuators. The interfaces of the

Figure 12: Integration and testing of ECUs with an In-the-Loop test system



test system need to be modified accordingly. Any intermediate steps can also be incorporated into the procedure.

A structure as illustrated in Figure 12 depicts the ECUs in the form of a black box. The behavior of the ECU functions can only be assessed on the basis of the input and output signals w, y and u^*. This "black-box view" is sufficient for simple software functions. But testing more complicated functions requires the integration of a measurement procedure for internal ECU intermediate variables. This type of measuring technique is also known as instrumentation. The testing of diagnostic functions also requires access to the fault memory via the ECU diagnostics interface, and this requires the integration of a measurement and diagnostics system.

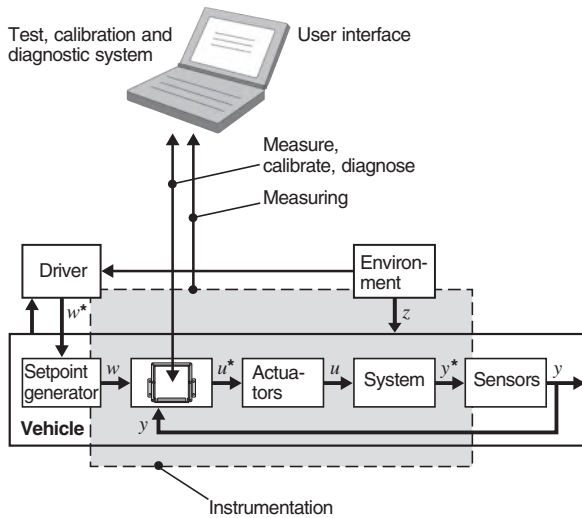
In-the-Loop test systems

In-the-Loop is a test procedure in which an embedded electronic system is connected via interfaces to a real (e.g. sensors, actuators) or virtual environment (mathematical models). The reaction of the system is analyzed and played back to the system. Depending on the type of test specimens, a distinction is made between the following test systems:

- In the case of Model-in-the-Loop (MiL), the function model of the software is tested. The model runs on a development computer.
- In the case of Software-in-the-Loop (SiL) the software code is tested. It runs on a development computer.
- In the case of Function-in-the-Loop (FiL), the software code is also tested. However, in contrast to SiL, this runs on the destination hardware. Coupling between the software and the environment model is performed by means of hooks and an emulator probe.
- In the case of Hardware-in-the-Loop (HiL), the complete ECU is tested by means of the input and output interfaces (I/O). Combinations of FiL and HiL are also used. PCs are increasingly being used as simulation computers.

In-the-Loop test systems can be used to validate and further develop software and hardware.

Figure 13: Integration and testing of ECUs in the real vehicle



Calibration of software functions

Procedure

Each electronically controlled vehicle system can only then develop its capability when it is optimally adapted to the respective vehicle type. In order for software functions to be used in as many vehicle variants as possible, these contain variable parameters. Adapting these parameters to the corresponding vehicle variant and for every operating condition (e.g. cold running, extreme heat or altitude) is known as calibration. To reach the desired level of functionality in the complete vehicle, a multitude of characteristic values, characteristic curves and program maps – the calibration data – must be applied.

Most changes to the vehicle require a change to the calibration. An example here: The λ oxygen sensor in the exhaust-gas system measures the residual oxygen in the exhaust gas. The fuel quantity actually injected can be determined on the basis of this signal. In this way, it is possible to adapt the activation parameters in the engine ECU exactly on a continuous basis. Changes to the exhaust-gas system which result in a change to the exhaust-gas back pressure at the point where the λ sensor is installed (e.g. altered exhaust manifold, particulate filter) must be adapted. If this is not done, this will result in a deterioration in consumption and emissions.

Variant management can be simplified in development, production and service by separating program and data statuses. The program status contains, for example, all the information on the variables to be applied and their limit values and correlations, while the data status contains the variables actually applied.

Calibration takes place in the laboratory on engine and vehicle test benches during vehicle tests and under real environmental conditions on test tracks. In addition to a measurement and diagnostic system, a calibration system is often required to calibrate internal ECU parameters (such as characteristic curves and program maps). On completion of calibration, the determined data are comprehensively checked. These values are then stored in the read-only memory (EPROM or flash) of the series ECU.

The parameter values must be variable during calibration. A calibration system, therefore, consists of one or more ECUs with a suitable interface to a measurement and calibration tool (Figure 14). In addition to their deployment in vehicles, measurement, calibration and diagnostic systems can also be used in in-the-loop test systems and on test benches. It is being applied more and more often in a virtual environment. Here, high-performance tools assist in finding the trade-off (optimal compromise of the calibration data).

Changes to the parameter values, for instance to the values of a characteristic curve, are supported in the calibration tool by editors. Alternatively, these operate on the implementation level (i.e. with the applied values) or on the physical specification level. Accordingly, the measurement tool converts recorded values into a physical representation or an implementation representation. Figure 14 shows an example of the physical and the implementation levels for a characteristic curve and a recorded measurement signal.

When working with calibration systems, it is usually possible to choose between offline and online calibration.

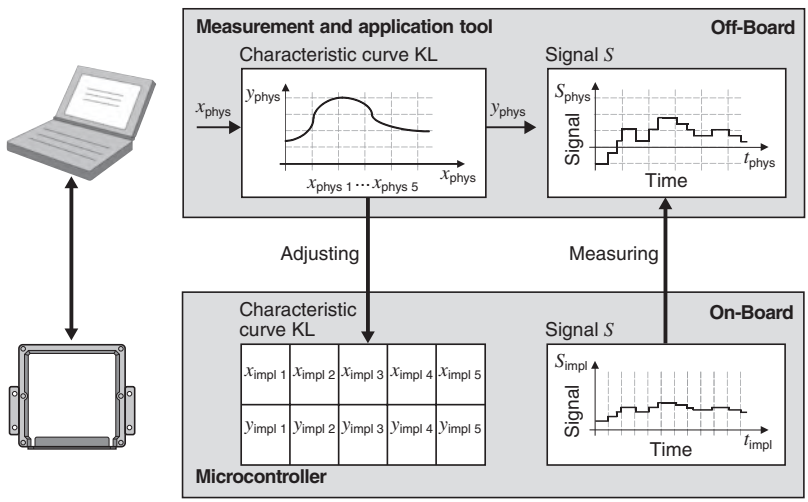
Offline calibration
In offline calibration, the execution of open-loop control, closed-loop control, and monitoring functions, i.e. of the "drive program", are interrupted to modify or adjust parameter values. Offline application therefore leads to many restrictions. In particular when deployed on test benches and during in-vehicle testing, it always causes an interruption to the test-bench or road test.

Online calibration
In online calibration, the parameter values can be adjusted while the microcontrollers are executing the software functions. It is, therefore, possible to adjust parameter values while executing open-loop control, closed-loop control, and monitoring functions at the same time, and hence during regular use at the test bench or in vehicle.

Online calibration places higher demands on the stability of the open-loop control, closed-loop control, and monitoring functions, since the driving schedule must remain stable during all adjustment procedures if exceptions occur, for example, if the distribution of interpolation points in characteristic curves fails to rise monotonically for short periods of time. Online calibration is suitable for long-term modification of less dynamic parameters (for example, tuning of engine control functions on the engine test bench).

For applying more dynamic functions or safety-critical functions (for example, setting the software functions of an ABS system for braking maneuvers during a test drive), the settings are not adjusted during actual braking maneuvers. In this case, online calibration can still save time by avoiding any interruptions in the program execution, thus reducing the interval between two test drives.

Figure 14: Operating concept with measurement and calibration tools
S Signal, t Time, x Input variable, y Output variable.
Indices:
phys Physical representation;
impl Implementation representation



Outlook

New vehicle functions and technologies will continue to lead to an increase in the scope of software in motor vehicles – even in the low price segment. In the process, the number of ECUs stays the same in many vehicles. As a consequence, the scope of functions of some ECUs is continuing to increase. In all, the system limits between the different ECUs are disappearing more and more, such as, for example, in the management of a hybrid drive. The electronics in motor vehicles are thus becoming ever more complex – including beyond an individual vehicle. The economic mastery of complexity is therefore certainly a key element in future automotive software engineering and the associated development environments.

Another trend is the increasing virtualization of development. Virtualization involves integrating parts of later development steps, such as testing and calibration, into the earlier development phases such that the function models can be optimally tested in-the-loop and fed with data on a preliminary basis early on. In this way, failures are detected earlier, support is provided for distributed development and bottlenecks shortly before series startup are avoided.

References

[1] Konrad Reif: Automobilelektronik – Eine Einführung für Ingenieure, 4th Edition, Vieweg+ Teubner Verlag, 2012.
[2] Jörg Schäuffele, Thomas Zurawka: Automotive Software Engineering – Grundlagen, Prozesse, Methoden und Werkzeuge, 5th Edition, Springer-Vieweg-Verlag, 2013.
[3] Werner Zimmermann, Ralf Schmidgall: Automobilelektronik – Bussysteme in der Fahrzeugtechnik: Protokolle und Standards, 4th Edition, Vieweg+ Teubner Verlag, 2011.
[4] Robert Bosch GmbH: Autoelektrik und Autoelektronik, 6th Edition, Vieweg+Teubner Verlag, 2010.
[5] ASAM website: http://www.asam.net/.
[6] IEC website: http://www.iec.ch/.
[7] ISO website: http://www.iso.org/.
[8] IEEE website: http://www.ieee.org/portal/site.
[9] MISRA website: http://www.misra.org.uk/.
[10] SAE website: http://www.sae.org/.
[11] OSEK VDX Portal website: http://www.osek-vdx.org/.
[12] AUTOSAR website: http://www.autosar.org/.
[13] ISO 14230: Road Vehicles – Diagnostic Systems – Keyword Protocol 2000, 1999.
[14] ISO 15765: Road Vehicles – Diagnostic Systems – Diagnostics on CAN, 2000.
[15] ISO 22900-1: Road vehicles – Modular vehicle communication interface (MVCI) – Part 1: Hardware design requirements.
[16] ISO 22901-1: Road vehicles – Open diagnostic data exchange (ODX) – Part 1: Data model specification.
[17] V model of IABG website: http://www.v-modell.iabg.de/.
[18] DIN EN ISO 9000: Quality management systems – basic principles and terms (ISO 9000:2005); Three-language version EN ISO 9000:2005.
[19] ISO/TS 16949: Quality management systems – Particular requirements for the application of ISO 9001:2008 for automotive production and relevant service part organizations.
[20] CMMI website: http://www.sei.cmu.edu/cmmi/.
[21] Automotive SPICE website: http://www.automotivespice.com/.
[22] DIN ISO/IEC 15504: Information technology – Process assessment –
Part 1: Concepts and vocabulary (ISO/IEC 15504-1:2004).
Part 2: Performing an assessment (ISO/IEC 15504-2:2003 + Cor. 1:2004).
[23] DIN EN 61508: Functional safety of electrical/electronic/programmable electronic safety-related systems – Part 1: General requirements (IEC 61508-1:2010); German version EN 61508-1:2010.
[24] ISO 26262: Road vehicles – Functional safety.

Architecture of electronic systems

General

Owing to the increasing demand for safety, comfort, entertainment, and environmental protection, electronic systems in motor vehicles realize an increasing number of functions and are characterized by a high level of networking and complexity. Optimized processes, methods, and tools of system architecture are required to keep on top of this structure in the future too.

History

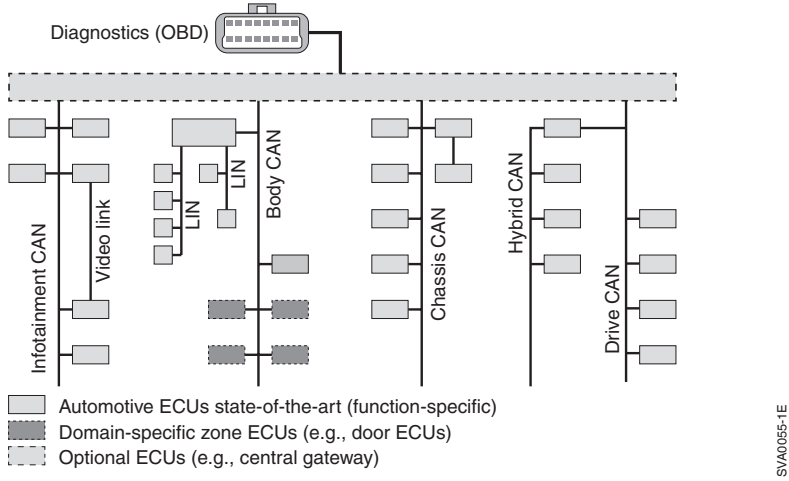
Over the many decades of automobile history, there has been a manageable number of electrical systems in motor vehicles: ignition, lighting, windshield wipers, horn, fuel gage, various indicator lamps, and a vehicle radio. Semiconductors were used initially – except in vehicle radios – only for rectification (direct-current generator replaced by alternator from approximately 1963) and then later for electronic control (transistorized ignition from 1965).

Certain in-vehicle functions were realizable with electromechanical means or with discrete electronic components either not at all or only with disproportionately high complexity. Thus, for example, the first electronic antilock braking system (ABS) had already been developed in 1970, but was never ready for series production on account of its size, weight and cost. By the mid-1970s the development of integrated circuits for a broad range of applications had also reached and revolutionized automotive engineering.

One of the first instances of the networking of electronic systems came about during the development of the traction control system (TCS). This networking was initially realized by purely mechanical means. The throttle valve in the air-intake system of the internal-combustion engine was fitted with a device which could be activated directly by the traction control system. It was not discernible to the engine management whether the driver or the traction control system was moving the throttle valve.

The next stage involved the realization of an electronic connection to the engine control unit via a PWM interface (pulse-width modulation) to improve dynamic response. This could be used to transfer

Figure 1: ECU interconnection in a modern mid-size vehicle



the signal to the engine control unit for reducing the drive torque. This was then implemented in the form of an air-supply throttling, an injection blank-out or an ignition- angle adjustment.

On account of the ever more stringent exhaust-emission regulations the possibilities represented up until that point of coupling between the traction control system and the engine management were no longer sufficient. What was now required was to transmit to the engine management how a reduction of the drive torque requested by the traction control system is effected in the air, fuel or ignition path. It was therefore necessary to come up with a more powerful interface via which a desired torque and a dynamic-response request could be transmitted from the traction control system to the engine management. By contrast, the actual torque, the engine speed and the current setting reserve were to be transmitted to the TCS control unit. It proved complex and expensive in terms of the number of cables required to transfer these different data via discrete and, for example, pulse-width-modulated interfaces. The CAN bus system (Controller Area Network) was introduced in 1991 as an alternative to discrete cabling. In this way, the foundations for the modern networking of systems in motor vehicles were laid.

Technology of the present day

In today's vehicles, virtually all the ECUs are networked directly or indirectly (e.g. via gateways) with each other (Figure 1).Gateways facilitate the exchange of data between different communication systems and beyond vehicle limits (connection to radio systems and the Internet). Networking goes to some extent so far that 60 or more ECUs communicate via several CAN or CAN FD bus systems (CAN with Flexible Data-Rate) and further communication systems, such as FlexRay, MOST (Media Oriented Systems Transport) or LIN (Local Interconnect Network), with each other. Aside from the classic communication systems, Ethernet is also used for various vehicle functions which require a large bandwidth. This is the case for example with a 360° camera for automatic park assist, which enables the driver among other things to have a bird's-eye view of the vehicle for example in the infotainment-system display.

Because of the powerful networking between the ECUs, some new performance features can even be achieved completely without additional hardware, i.e. purely by means of data communication and software. An example of this is the tire-pressure monitoring system (TPMS). Since the end of 2014 a tire-pressure monitoring system has been mandatory in all vehicles. There are two ways of realizing such a system: firstly as an indirect solution and secondly as a direct solution. The indirect solution uses the wheel-speed sensors of driving-dynamics control (Electronic Stability Control, ESC) and thus does not require any further hardware. As soon as a wheel loses air, the wheel diameter decreases, as a result of which the wheel must turn more quickly and consequently the air loss can be determined by means of an algorithm.

Also classed as state-of-the-art are numerous Advanced Driver Assistance Systems (ADAS). For example, the car manufacturers offer functions such as Adaptive Cruise Control (ACC), Lane Keeping Support, or the combination of adaptive cruise control and lane keeping support known as Integrated Cruise Assist (ICA), and many others. All this can be realized with a matching E/E architecture with suitable sensors, such as for example a video camera, radar sensors, and ultrasonic sensors.

Development trends

The increasing number of electronic and electrical components in modern vehicles and the constantly increasing complexity necessitate a continuous further development and constant technological advances in E/E architectures. For this reason, E/E architectures of vehicles have in the past obviously changed with regard to the communication technologies and also in their systemic manifestation. Further developments will also be constantly identifiable in the future. The following text will examine the evolution of E/E architectures and provide an outlook of future developments and trends (Figure 2).

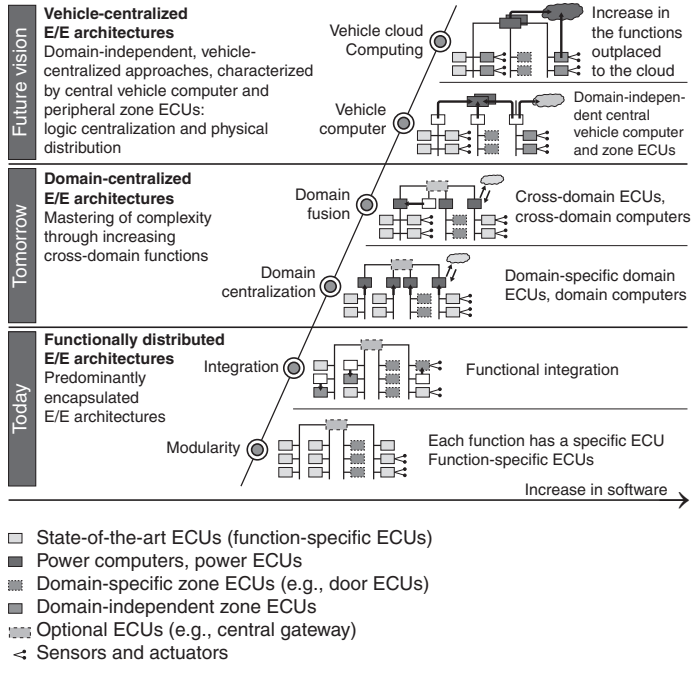
Distributed E/E architectures – modularity and integration
The picture of the first E/E architectures and their comparably lower level of complexity was initially characterized by their strong modularity (Figure 2, bottom third). Logic functions were distributed to many individual ECUs – function-specific ECUs – in the vehicle such that virtually every function had its own ECU with performed the function. In the ensuing period, with the sharp increase in functions, new functions were frequently integrated in ECUs that already existed in the vehicle. The reasons for this in particular are the rapidly growing number of ECUs and the associated rising costs. An example of functional integration is the integration of a parking-assist function in the "Body Control Module", which is essentially responsible for controlling and monitoring the functions and components of the body electronics (e.g. light-actuator technology, door-actuator technology).

For the reasons mentioned this stage of development of the E/E architectures is called "Functionally distributed E/E architectures" (Figure 2).

Domain-centralized E/E architectures
In the ensuing step – the domain-centralized and cross-domain-centralized E/E architectures – not only do the number of electronic and electrical components and the number of functions continue to rise sharply, but also increasingly the dependencies of the functions within the individual domains (Figure 2, middle third). The term "domain" refers to the grouping of similar functions. Typical domains are shown in Figure 3.

Furthermore, the interfaces and dependencies between the domains increase sharply. This development is already

Figure 2: Bosch E/E architecture roadmap



clearly discernible in today's vehicle architectures and will take on an ever more significant role in future. An example of this is the cross-domain function "Vehicle Motion Control", which will be described in more detail later. Thus, in this case, sensors, actuators, and functions from the "Powertrain" and "Chassis and Safety" domains interact directly with each other.

In other words, increasingly vehicle-wide functions extending over a number of domains stand out. The associated strong logic interaction and networking as well as the resulting dependencies give rise from a functional perspective to a centralization. This means that, instead of the originally modularly distributed arrangements, increasingly functions are being bundled into domain ECUs, domain computers, or event into cross-domain ECUs or cross-domain computers.

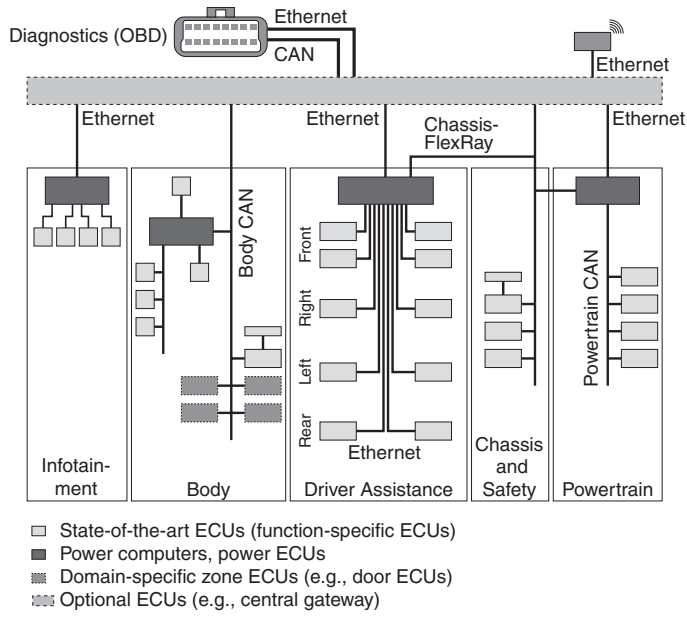
Therefore the term used for the E/E architectures of tomorrow is "Functionally domain-centralized E/E architectures" (Figure 3).

Vehicle-centralized E/E architectures

If one takes a further look into the future, the vehicle is in particular characterized by a high degree of automation, networked systems – vehicle-internal and vehicle-external – and by multimedia and infotainment applications. These vehicle properties impose the highest demands on the E/E architecture and on the communication network. In this way, requirements with regard to maximum permissible latencies or the number of hops (path from one network node to the next) acquire ever greater importance. This constitutes a huge challenge for classic E/E architectures. For this reason, the trend is towards functional centralization (Figure 2, top third).

If all the logic were to be brought together on a powerful central vehicle computer, an ideal state would be achieved for example with regard to the mentioned latency requirements and the number of hops. All the functional interfaces would then be located within this central vehicle

Figure 3: Schematic diagram of a functionally domain-centralized E/E architecture



computer. This is the basic idea behind centralized E/E architectures, consisting of a central vehicle computer which with extremely high computing capacity constitutes a vehicle-wide functional control system and a number of zone ECUs. While a large part of the complex logic of the highly automated and comfortable vehicle of the future will therefore be realized in the central vehicle computer, the peripheral zone ECUs function from a communication standpoint as gateways and from an energy standpoint as supplying and driving elements for the vehicle sensor and actuator technologies. In order to guarantee the high communication demand and hence the high bandwidths, technologies such as Ethernet will in future take up greater shares of the communication network and entail new concepts (Figure 4).

Furthermore, even outplacements of logic elements to the back end – the so-called cloud – are conceivable in order to be able to calculate cross-vehicle information and algorithms, for example the calculation of traffic forecasts, not locally but centrally and to make available a number of vehicles or vehicle fleets.

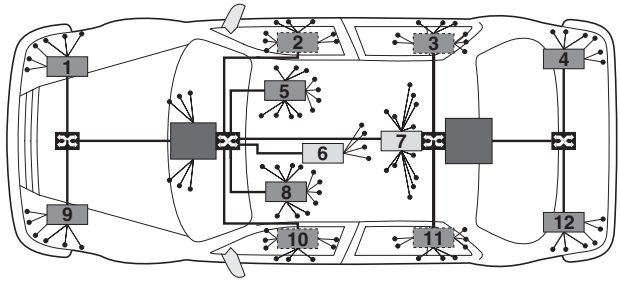
The following section describes using highly automated driving functions how these bring an influence to bear on the E/E architectures in modern vehicles.

Effect of highly automated driving functions on the E/E architecture

Nowadays car manufacturers increasingly advertise with driver-assistance functions which make it easier for drivers to negotiate road traffic safely and comfortably. As good as all the functions that can be bought today by the end customer are assistance functions in which the driver is supported but not replaced. This is to change in the near future. Car manufacturers and suppliers are meticulously developing highly automated driving functions to further relieve the load on drivers. A distinction is made here between assisted (driver has to monitor) and piloted (driver does not have to monitor) driving. The German car industry has in a panel of experts come up with a uniform understanding of the stages of automation [1].

Figure 4: Schematic diagram of a zone-orientated E/E architecture

1 Zone, front right
2 Zone, door, front right
3 Zone, door, rear right
4 Zone, rear right
5 Zone, front passenger
6 Zone, roof, right
7 Zone, rear seat
8 Zone, driver
9 Zone, front left
10 Zone, door, front left
11 Zone, door, rear left
12 Zone, rear left



Switch
Ethernet
Sensors or actuators
State-of-the-art Automotive ECUs (function-specific)
Central vehicle computer
Domain-specific zone ECUs (e.g., door ECUs)
Domain-independent zone ECUs

SVA0058-2E

Starting at Stage 0, in which the driver exclusively assumes longitudinal and lateral control of the vehicle, right up to Stage 5, in which the system can automatically manage all situations during driving and a driver is no longer required (Table 1). The factors driving the development of highly automated driving are, aside from greater road safety, above all increasing comfort and convenience and more time which the driver can use for other activities. Activities like for example writing e-mails, surfing the Internet, and many more are then possible while driving.

For highly automated driving to be possible at all, it needs a powerful and reliable E/E architecture. The following section will address which E/E architecture blocks and which adaptations are required for this.

E/E architecture blocks

To realize a powerful and reliable E/E architecture, certain E/E architecture blocks are needed. These include among others communication, the vehicle electrical system, the sensor and actuator set, and performant ECUs. For highly automated driving functions it is important that parts of these E/E architecture blocks are always available. This gives rise to high safety requirements, which can only be satisfied by fallback levels (redundancies) in the event of a fault. Such concepts are also called fault-tolerant. Here, redundancy can refer to communication, the vehicle electrical system, and the software and the hardware.

The E/E architecture blocks and corresponding fallback levels are described in the following.

Fault-tolerant communication network

Highly automated driving functions in vehicles must ensure that required information is always available between two components. For example, it must always be possible for the brake and the steering to receive the necessary information as to whether the vehicle hast to be braked or steered. Redundant communication can be used to satisfy this safety requirement. A distinction is made here between a

Table 1: Degree of automation acc. to VDA (source: VDA)

Stage	Degree of automation	Description	
		Driver	Automation
0	Only driver	Driver constantly performs longitudinal and lateral control	No intervening vehicle system active
1	Assisted	Driver constantly performs longitudinal or lateral control	System assumes the other function in each case
2	Partially automated	Driver must constantly monitor the system	System assumes longitudinal and lateral control in a specific application
3	Highly automated	Driver must constantly monitor the system Driver must potentially be in a position to take over	System assumes longitudinal and lateral control in a specific application System identifies system limits and prompts the driver to take over with a sufficient time reserve
4	Fully automated	No driver required in the specific application	System can automatically manage all situations in the specific application
5	Driverless		System can automatically manage all situations throughout the entire journey No driver required

variety of topologies: ring, star, bus, and fully meshed topology (Figure 5, see also Networking in motor vehicles).

Fault-tolerant vehicle electrical system
Like the communication network, the vehicle electrical system must also have a fault-tolerant design. The vehicle electrical system can be divided into three elements (Figure 6): storage elements (e.g. battery), coupling elements (e.g. DC/DC converter), and source elements (e.g. alternator).

To enable the vehicle to still be controlled accordingly in the event of a fault, the safety-related components, which for example implement the braking and steering functions, are supplied by two completely independent storage devices. These two storage devices are in the event of a fault separated via a coupling element, meaning that the fault is encapsulated and consequently the fault is prevented from spreading to the other supply channel. On account of the redundant actuator technology, in this case the safety-related functions like braking and steering continue to be available – for a limited time at least – for a transition to a safe state, even if a supply channel fails.

Sensor redundancy concept
An important influencing factor for highly automated driving is the vehicle environment. This is detected by a variety of sensors such as radar, lidar, camera, and ultrasonic sensors. GPS and inertial sensors are also used to pinpoint the vehicle position. In combination it is possible to determine both what is situated around the vehicle and where the vehicle itself is situated.

A redundant sensor set can also be realized with the basis of the redundant communication network and the redundant vehicle electrical system. It is import-

Figure 5: Fault-tolerant communication: ring, star, bus, and fully meshed topologies
a) Ring topology,
b) Fully meshed topology,
c) Bus topology,
d) Star topology.

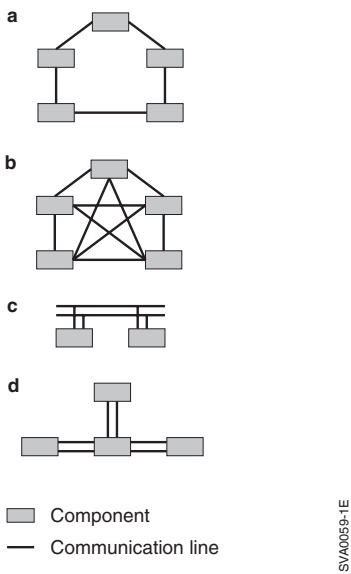
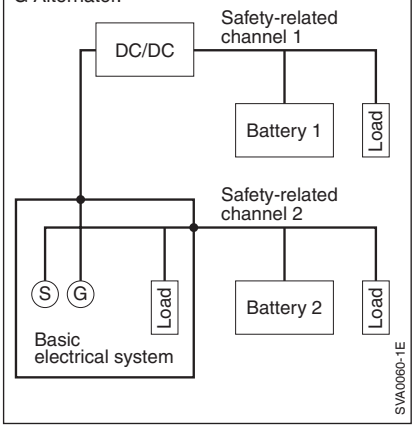


Figure 6: Fault-tolerant vehicle electrical system
S Starter,
G Alternator.



ant here that a field of view (FoV) is always covered by different sensor principles so that the environment can be reliably detected in all situations. For example, the sensors should be split between the two supply channels in such a way that the field of view is covered during normal operation by sensors A, B, C, and D, in fault scenario I (fault on supply channel 1) by sensors B and C, and in fault scenario II (fault on supply channel 2) by sensors A and D. Which sensor principles, sensor combinations, and sensor distributions are ideal to be able to detect the environment in all situations is currently the object of development.

Performant ECUs

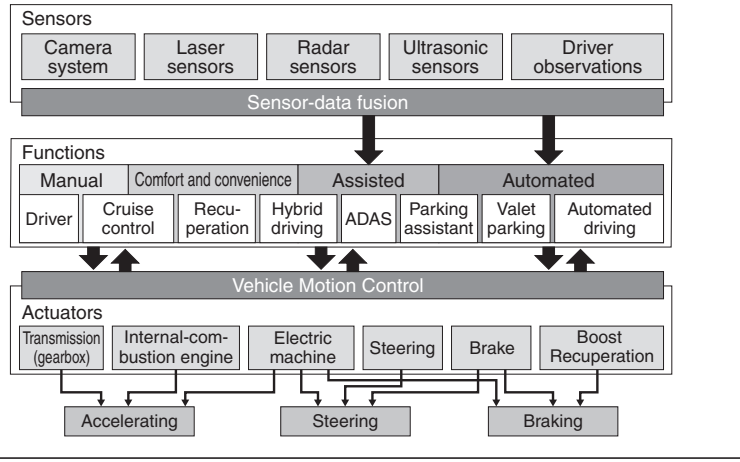
The vehicle environment and the vehicle position are elementary information for highly automated driving. The sensor data obtained are fused in performant ECUs into an environment model. A lot of computing capacity is required for this processing operation. Fault tolerance is important here so that the sensor-data fusion can also be computed in the event of a fault. Among others, for example, either a fault-tolerant ECU (e.g. two boards in one housing) or a combination of two ECUs with one main path and one substitute path can be used.

To be able to deliver this computing capacity at all, automotive electronics will in future increasingly turn to other disciplines, for example consumer electronics. This will also involve the use of microprocessors in addition to automotive-certified microcontrollers.

Vehicle Motion Control

Irrespective of whether assisted or highly automated driving is now employed when the driver wants to go from A to B, a "Vehicle Motion Control" (VMC) [2] (Figure 7) is always demanded. This VMC assumes

Figure 7: Vehicle Motion Control (source: [2])



lateral and longitudinal control of the vehicle without permanent monitoring by the driver by coordinating and monitoring access to actuators (brake, steering, and drive). The environment model from the sensor-data fusion serves as the input for deriving an adequate driving decision.

Actuator redundancy concept
Highly automated driving functions such as highway pilot require a redundant brake and steering in the network of actuators which are coordinated by VMC. Redundant braking capability is provided here for example by the combination of a vacuum-independent, electromechanical brake booster and driving-dynamics control (ESC) (source [2]). Furthermore, fault-tolerant steerability for example is provided by electric steering in which all the components are redundantly designed (among others power supply, communication, windings, logic, etc.).

E/E architectures for highly automated driving
For the various highly automated driving functions such as for example traffic-jam pilot or highway pilot the previously mentioned E/E architecture blocks are required to realize a powerful and reliable E/E architecture. The described stages of the degree of automation can be used to subdivide the driving functions and derive appropriate requirements with regard to the E/E architectures. Reference is made here once again to the importance of the redundant design of the individual E/E architecture blocks for highly automated driving functions which do not require monitoring by the driver.

Development of E/E architectures

Definition of "E/E architecture"
As the amount of electronics and networking in the vehicle increases, so too the demand for powerful development processes and their description methods for the architecture of electrical and electronic systems increases.

The term "architecture" generally refers to the art of building. In the construction industry, the architect designs a building by drawing up plans for the different views and contractual work based on the client's wishes and boundary conditions. A plan abstracts the reality with regard to a particular aspect (e.g. geometric conditions or electric cabling). The building can finally be erected on the basis of the plans of all the necessary aspects.

When carried over to a motor vehicle, this is referred to as the "E/E architecture". "E/E" denotes the electrical and electronic aspects of the motor vehicle. The "plans" of the E/E architect are referred to in the following with the general term of "model".

Car manufacturers and their suppliers have different views on how many models of which type are needed to describe completely the electrical and electronic systems in the vehicle. The models presented in the following have proven successful in practice and are a necessary framework for describing the E/E scope.

The term architecture is often used in the literature and in publications to denote the models themselves. Here, a clear distinction is made between work operation (architecture development) and presentation of the result (model).

Models of E/E architecture
The models of E/E architecture reflect the results of the different integration aspects of the electronic systems in the vehicle (Figure 8). These aspects are usually dealt with simultaneously because E/E architecture development in the concept phase of a vehicle addresses both the geometry (the "body structure") and new systems which are to be implemented in the vehicle. In the course of vehicle development, the situation may for instance arise where an electronic system in the

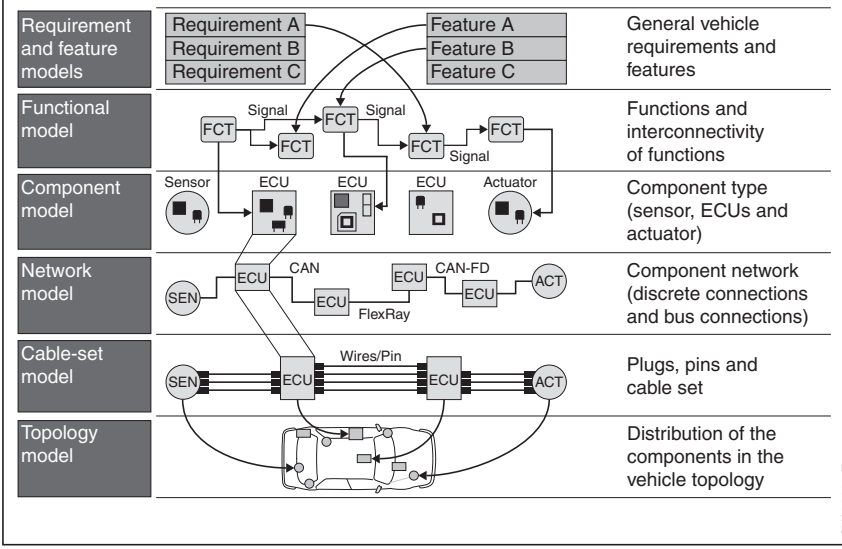
chosen technology cannot be integrated into the available space. Compromises must be found in this case.

Feature and requirement models
Feature model
The feature model defines which performance features (features, fixed term in E/E architecture) the vehicle is to have later. These performance features are vehicle properties which can be experienced by the driver and the passengers – thus for example ACC Stop and Go (see Adaptive Cruise Control). The vehicle scope is defined with the feature model. All further models can be derived and developed on the basis of the features.

Requirement model
All the necessary vehicle requirements are set out in the requirement model. These put the features in concrete terms and serve as an input variable for the functional model. Here, for example, the ACC feature can be put in concrete terms with regard to the driving speed. Because the Stop-and-Go variant of ACC is involved, the lower limit of the speed range must be 0 km/h. The upper limit differs from car manufacturer to car manufacturer (e.g. 130 km/h). If the feature for example were not a Stop-and-Go ACC variant, the speed range could be 30 to 200 km/h.

Functional model
Functional models are the preliminary stage of concrete technical systems. They describe the logic functions which are needed to realize the required performance features without going into their concrete technology. Thus the functional architecture is independent of the technical architecture. It is thereby possible to develop features optimally and without technical restrictions. For the example of ACC Stop-and-Go (Figure 9) this entails a breakdown into logic functions such as:
– Sensor technology,
– Longitudinal control of braking deceleration,
– Longitudinal control of acceleration,
– Actuator technology.

Figure 8: Models of E/E architecture
FCT Function, ECU Electronic Control Unit, SEN Sensor, ACT Actuator.



Component model

Technology model

The technology model describes what technical realization is used for the specified transfer elements without already combining these into modules, such as electronic control units (ECUs). "Technology blocks" are created.

Thus, signal filtering can be realized with discrete components by means of a digital circuit or filter software on a microcontroller. Even a controller function can be executed with a discrete electronic circuit or a microcontroller. Voltage stabilization can be achieved by either a smoothing capacitor or a DC/DC converter.

The decision as to a technical implementation is influenced not only by the function, but also by further criteria such as costs, weight, scalability, etc. Before the technology blocks are combined into modules in the form of ECUs, the first step is to look for synergy with the further technology blocks to be integrated. A technological active chain is created (Figure 10). If a specific sensor technology is available, for example, for an active-chain link whose signal is required by another active chain, this will also be used. This occurs even if this sensor is overspecified for the additional user, i.e. there are lower requirements, for example, of the available signal range or of the accuracy.

It is nevertheless important to store the original requirement in a database as this synergy may no longer be present in another vehicle. The automotive industry usually uses the nomenclature according to DIN EN 60617 [3] to describe the hardware.

Node model

The links of the technological active chains are combined into groups at different locations, the "nodes". Here, strict adherence to the optimum cost and further criteria (e.g. latencies) is maintained in relation to integrating the technology blocks. Thus, attempts are made, for example, to integrate the software parts of several technological active chains on a common microcontroller. Sensor information is used jointly wherever possible and actuators are activated by different functions. For example the ACC and ESC functions activate the brake actuator.

Figure 9: Functional model ACC

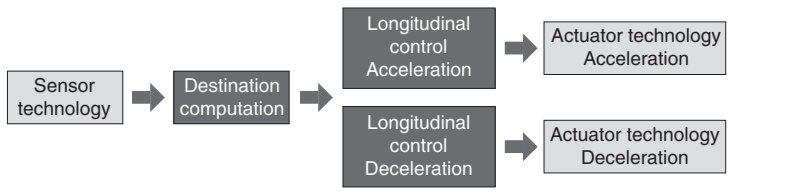
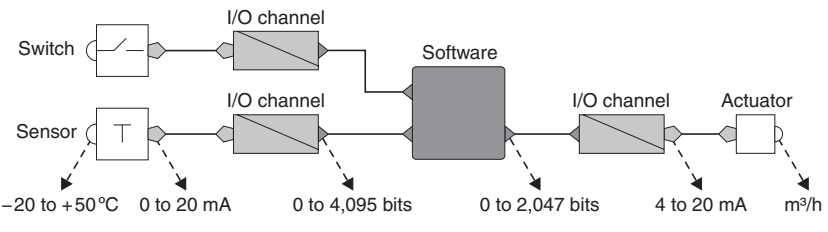


Figure 10: Example of a technological active chain



Hardware model of the electronic component
This model represents the structure of the electronic hardware of an individual ECU. It is created by allocating specific electronic components from the technological active chains to an electronic module in a node. An ECU is therefore, generally speaking, a collecting point for electronic components of different systems, an "integration platform".

Software for controlling different systems from different sources (car manufacturers or their suppliers) is integrated on the microcontrollers located in the ECU. By networking the ECUs, it is possible to implement complex distributed functions which utilize sensors and actuators from different installation locations in the vehicle.

During development, initially customary circuit diagrams will be used for the electrical or electronic parts of the ECU. Then the ECU mechanics and the design and connection technology will be established. E/E-architecture development is confined to a very rough representation in the early concept phase.

Software model of the electronic component
The AUTOSAR standard defines the structuring of software close to hardware level and its interface to the application functions and establishes interfaces between application functions. AUTOSAR additionally defines standardized exchange formats which are supported by popular modeling tools (see AUTOSAR).

Typically, a distinction is made between basic and application software. Blocks of basic software are, for example, device-driver software, communication software, operating system, and hardware abstraction.

Network model
Network model of communication
Because all the technology blocks of a vehicle have been allocated to ECUs in the previous steps, a network of these ECUs is now in place with their communication relationships. The network model of communication represents all the ECUs in the motor vehicle which have bus communication and are thus directly or indirectly networked with each other (see for example Figure 1).

Each signal which is exchanged between two or more ECUs is assigned to a suitable bus system. AUTOSAR defines for this purpose standardized exchange formats which enable bus communication to be described.

Network model of energy supply
The allocation of the technology blocks to ECUs and sensor and actuator modules has also given rise to a network of electrical loads/consumers which requires a suitable energy supply. On the one hand, it is important to fuse individual electric circuits so that a short-circuit does not affect the entire network. On the other hand, not all circuits should be supplied with electrical energy in each operating state. The principle of "terminals" was therefore introduced for this purpose. Thus, for example, terminal 15 is only supplied with electrical energy when the ignition is switched on. The assignment of the individual terminals to ECUs follows the "terminal concept". Among others, the different vehicle states (e.g. driving, parking, loading, etc.) are required as input for the terminal concept. For example, parts of the energy management and the wake-up and go-to-sleep sequences of the ECUs are realized in the network by means of hardware.

The electrical circuit diagram (Figure 11) shows the electrical networking and fusing of the individual modules without the installation position being taken into consideration. Here the colors of the cables (not shown in the figure) and the matching with a terminal or fuse can be seen. The terminal designations follow the conventions in DIN 72552 [4].

The positive pole of the supply voltage is usually featured in the top half of the representation while the negative pole (ground) is featured in the bottom half.

Cable-set and topology models
This model groups electrical and electronic modules in a specific location in the vehicle (Figure 12). In this way, the connecting cables between the ECUs and the energy-supply leads of the electrical

loads/consumers are brought together in cable looms. This creates the cable set. Here many different boundary conditions must be observed, such as for example:
– the manufacturing concept (one-part or multi-part cable set),
– the loom cross-sections (flexibility),
– the electromagnetic compatibility (EMC),
– the heat dissipation,
– the weight,
– the costs (e.g. of copper),
– and the structure of the cable set in the vehicle.

The topology describes the possible routing paths in the body, such as for example the H structure, which consists of two main connections from the front end to the rear end of the vehicle and a cross cable (crossing) from the left to the right side of the vehicle.

Two-dimensional models are usually sufficient in the concept phase of a vehicle; detailed three-dimensional models are used in the later development phase.

E/E architecture development process
The E/E architecture development process links the individual draft stages with each other on a logic and time basis and provides quality criteria at the beginning and the end of a draft stage. The individual draft stages are described below.

Requirement management
The requirements decisively determine the decisions of the E/E architect. It is advisable and common to distinguish between functional and non-functional requirements. Functional requirements refer to the desired performance features, the features, which can be used in the vehicle. Non-functional requirements refer to the technical solution and are therefore also known as draft restrictions.

Such a restriction can be, for example, the space available in the center console for installing ECUs. Another restriction can be the maximum permissible heat dissipation in a location which influences the power electronics positioned there.

Figure 11: Circuit diagram using example of a vehicle radio
15, 30, 31 Terminal designations,
A2 Vehicle radio, W1 Vehicle antenna, F Fuse,
B11, B12 Speakers, P6 Timer,
X18 Diagnosis socket,
1 to 8 Section identification.

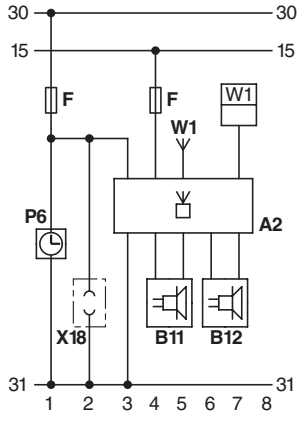
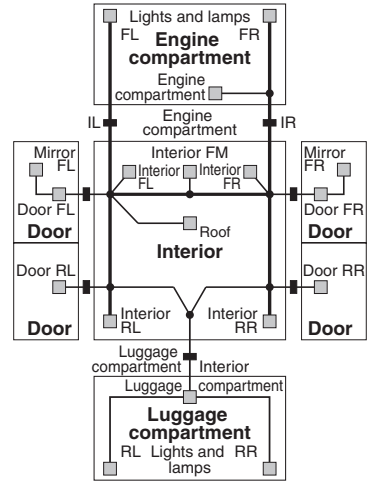


Figure 12: Example of a two-dimensional space model
RL rear left, RR rear right,
FL front left, FR front right,
FM front middle, IL inside left,
IR inside right.



Thus, for example, the audio amplifier in vehicles is frequently installed in the luggage-compartment area since the heat in the vehicle passenger compartment cannot be adequately dissipated.

Once the documentation of the functional and non-functional requirements has been completed, the actual development of the E/E architectures begins.

Development of E/E architectures

The development of the E/E architecture can follow two paths: The bottom-up approach, which starts out from existing components, and the top-down approach, which features the implementation of all the previously described modeling steps starting out from the functional and non-functional requirements.

In the bottom-up approach the creation of the E/E architecture starts out from the functionality of existing components and this is supplemented additionally by function and communication aspects. This approach is typically chosen for the creation of E/E architectures of successor generations to existing vehicle platforms or if cost restrictions necessitate the use of cost-effective components already existing on the market.

The top-down approach focuses on the function complexity and is typically chosen for creating E/E architectures of new vehicle platforms. In addition, the functional model will in future play an increasingly more important role, particularly to be able to realize cross-domain features. The functional model will be developed independently of the further technical implementation of the E/E architecture.

The use of E/E concept tools allows an exchange of data with development partners for electronic components or for the cable set.

Evaluation of models

The following must be observed for all approaches: During the transition from one model hierarchy to the next (e.g. from the functional model to the technology model), a list of evaluation criteria (e.g. reuse or testability) is compared with a portfolio of specimen solutions (e.g. bus technologies). Evaluation of the specimen solutions using the criteria allows a solution to take shape on the basis of the purely functional requirements and irrefutable boundary conditions ("MUST criteria").

An alternative procedure consists in comparing a reference solution (e.g. the previous networking model) using the evaluation criteria with alternative solutions. This does indeed deliver fast results, but possibly not the global optimum.

Because the evaluation criteria are generally weighted differently by the car manufacturers, the electronic systems of the vehicles sometimes differ considerably from each other.

E/E architecture development tools

Ideally a tool is used for architecture modeling which can represent and network with each other the different models and model levels of the E/E architecture work. Fully connected documentation is created in this way. This enables the different disciplines involved in the development process to come on board at the relevant points. It should furthermore be possible to record the modeling properties numerically in order to be able to provide them with an evaluation.

Different E/E architecture development tools have in the meantime come onto the market which enable architecture modeling to be performed on a tool-assisted basis. An important point is the standardization of the models and their data formats. Only this facilitates competition between the tool manufacturers and opens up to the different disciplines involved the opportunity of joining the process at different points.

Summary and outlook

As a result of the ever increasing scope and the increasing networking of electronic systems, it is necessary to use suitable processes, methods and tools in E/E architecture development. E/E architecture development has established itself as an independent area of responsibility in the automotive industry and has a decisive influence in the development of new vehicles. It is thereby possible also to render further features in the automobile. These include for example highly automated driving functions like highway pilot or wireless software and firmware updates via the air interface ("over the air"). It is possible through new approaches in E/E architecture development to deliver powerful and reliable E/E architectures in the future as well. If the procedures described in this publication are consistently applied and developed further, the increasing complexity in motor vehicles will also be manageable in the future and will continue to make significant contributions to improving traffic flow, traffic safety, driving comfort, and economical fuel utilization.

References

[1] VDA - https://www.vda.de/de/themen/innovation-und-technik/automatisiertes-fahren/automatisiertes-fahren.html.
[2] Automated driving, electrification and connectivity – the evolution of vehicle motion control. Alexander Häussler, Robert Bosch GmbH (DOI 10.1007/978-3-658-09711-0_3).
[3] DIN EN 60617: Graphical symbols for diagrams.
[4] DIN 72552: Terminal markings for motor vehicles.

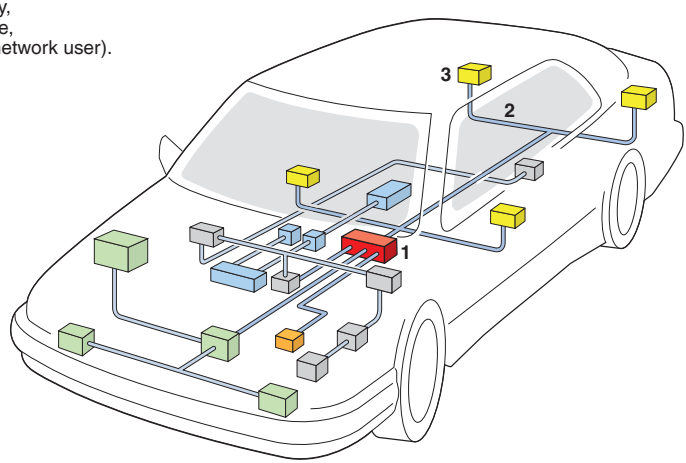
Vehicle communication networks

Bus systems

Networks for data communication, also known as bus systems or protocols, are widely used in today's motor vehicles. Various components such as sensors, actuators or ECUs – the nodes or network users – are connected to each other via a single channel (Figure 1). A wealth of data, also known as messages, telegrams, packets or frames, is exchanged via this channel. For example, the driving speed is determined in the Electronic Stability Program (ESC) and transferred to all the other ECUs which are networked as nodes in the bus system.

Advantages of bus systems
Compared with conventional cabling, in which the transmitter and receiver of information are each connected by separate lines, bus systems offer significant advantages:
– The costs of materials for the cables are lower (which offsets the higher costs for the electronics).
– The space required and the weight of the cabling are lower.
– The number of plugs susceptible to faults is lower, thus fewer failures occur in total.
– Data can be distributed to various receivers; for example, the signals of a sensor can be used by several systems.
– All the systems in the vehicle which are connected via the bus can be reached from one access point. This provides for simpler diagnostics and the configuration of all the ECUs at the end of the line.
– The performance of calculations can be distributed to different ECUs.
– Analog sensor signals must be digitized for data processing. The sensor signals can be conditioned directly in the sensor, the information is then disseminated via the bus.

Figure 1: Automotive networking
Schematic representation of nodes and data lines.
1 Gateway,
2 Data line,
3 Node (network user).



Requirements of buses

General requirements

In order to be able to be used in vehicles, buses must satisfy typical requirements. A bus must use a transmission process which has control over the propagation times, attenuations and reflections of the signals for cables up to 40 m long with which typically all the parts of the vehicle are reached. A network must be able to connect several dozen users in the process.

A bus must be able to withstand the rough ambient conditions in the vehicle with regard to temperature, vibrations and electromagnetic interference.

Because vehicles are produced in high numbers, even small savings potentials in the cost of the bus hardware of the individual vehicle must be utilized in the design of the bus. It must equally be noted that there are several competing manufacturers of bus components.

There is a large number of equipment variants of a vehicle type which must be covered by a basic configuration of the bus. The additional installation of an optional extra must not affect the other vehicle systems.

The performance of the bus must be clearly defined in a public standard in order to have a benchmark against which components can be verified. This ensures that verified components of different suppliers function together in a single network.

Special requirements

Different systems in the motor vehicle have different requirements and therefore necessitate the use of different buses. The required data transfer rate (data rate) very much depends on the application. It ranges from a few bit/s for switching the lights, through engine management with a few 100 kbit/s, right through to video applications with several Mbit/s.

For systems in which a failure or a delayed execution of a function is safety-relevant (e.g., airbag, electric power steering), a maximum duration for the delay of data transfer – the length of time between the sending of a telegram and actual reception (latency) – must be guaranteed in all cases. When the time response of the bus is defined and reproducible at all times, this is referred to as determinism. In this way, transfer times for messages, in particular, are already known in the design phase.

For safety systems such as electric power steering it must be verified that everything has been done in terms of state-of-the-art technology to avoid flaws/faults in design and to ensure correct functioning under all permissible boundary conditions.

Transfer malfunctions which inevitably occur during operation must be detected and dealt with for the relevant application.

Technical principles

Components of a bus

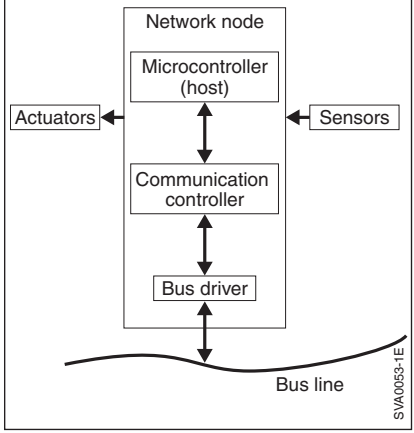
In order for a network to function as planned, calculations must be carried out constantly in every node. This is performed for the most part by a special piece of hardware (Figure 2), the "communication controller". This removes the load from the actual computer (host), which can also perform this function – although less efficiently. The communication controller can be implemented as a separate semiconductor component; in the case of many microcontrollers, however, communication controllers for some buses are already integrated.

A further component, the bus driver or transceiver, converts the signal which it receives from the communication controller into the physical signal on the bus line (e.g., voltage level). It converts data received on the bus line and forwards them to the communication controller.

Communication protocol

Data transfer between the nodes is established in the communication protocol. These agreements relate to the syntax, semantics, and synchronization of communication. The syntax determines which character strings may be used for communication, the semantics define the correct use of syntactically correct constructs.

Figure 2: Hardware of a node (network user)



OSI reference model

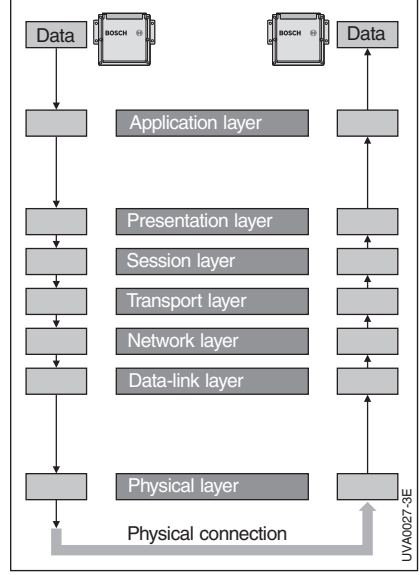
The ISO (International Standardization Organization) developed the OSI reference model (Open Systems Interconnection), which is often used as the basis for describing and comparing communications protocols. In this model, the function of a data-communication system is broken down into different hierarchical layers which each use the functions made available by a different layer. The OSI reference model is an aid to structuring conceptually the function of a communication system. However, finding efficient solutions is not necessarily supported by observing this model.

Seven layers are defined (Figure 3), but many protocols are only defined on one part thereof. The upper layers in the OSI reference model (application layer, presentation layer and session layer) are for the most part not served by motor-vehicle buses.

Physical layer

The physical layerdefines the physical properties of the transmission medium (e.g., voltage level or shape of plugs). In

Figure 3: OSI reference model



motor vehicles for the most part, electromagnetic signals in the kHz to MHz range are used on special cables and optical fibers (physical connection, Figure 3). Radio of different frequencies or the joint use of cables already provided with electrical power for supply purposes are technologies which have not yet been tested in large batches.

The most widely used method is, because of the low costs involved, transmission by cable, particularly as a voltage difference on twisted two-wire cables, with or without shielding against radiation, or on single-wire cables with a voltage reference to ground.

Optical fibers made from plastic or glass fibers (for the infrared spectral range) are used above all wherever high data rates are needed. They are insensitive to electromagnetic radiation (e.g., by the ignition system), but are costly to lay and their resistance to aging has not yet been adequately assured.

The physical possibilities of the medium also limit the types of encoding of a bit. With optical media, there are the two states “light” and “no light”, which offer encoding by means of amplitude (brightness).

Electrical voltage signals on a cable offer different possibilities for representing bits. The simplest and usual method used in motor-vehicle buses is encoding in which each bit is assigned a voltage value (amplitude modulation) which is applied for the entire bit duration (NRZ, Non Return to Zero). In contrast, with Manchester encoding the level changes during a bit length from 0 to 1 (corresponding to the encoding of 0) or from 1 to 0 (corresponding to the encoding of 0).

Because the signal level can remain constant over several bit lengths with NRZ encoding, additional measures are required to safeguard synchronization. With bit stuffing a complementary bit is inserted after a defined number of equal bit values; the receiver filters this bit out again.

In the case of bus systems with a high data rate, open cable ends must, if necessary, be terminated by terminating resistors in order to prevent reflections from disrupting the signal shape.

Data-link layer
The data-link layer effects the correct transport of data between neighboring nodes. Data bits are arranged into frames. By adding further bits such as checksums or numberings, it is possible to detect or even correct errors that occur during transfer. Alternatively, an error can be rectified by requesting a new transfer.

Network layer
If not every node is directly connected with every other, a route must be found for the data which passes through intermediate stations. The search for a route is also called routing and includes different characteristics of the transmission paths (particularly bandwidth) and the intermediate stations, also called routers.

Transport layer
The transport layer’s functions include disassembling large data packets and reassembling them at the receiver, which the parts potentially reach via different routes at different times, or if transmission errors occur ensuring that a packet is transmitted again. The transport layer also serves to hide the properties of the transmission path from the application and to offer services independently.

Data frame
Data are usually transmitted as fixed-structure packets (frames), which, as well as the actual content – the user data – also contain control and test data such as source, destination, priority and check bits to protect against corruption.

Data rate
The data rate is a measure of the speed at which the individual bits are transferred. It is given as the number of bits per second. This produces the derived units such as kbit/s and MByte/s (1 byte = 8 bits).

The gross data rate takes into account all the transferred bits, while the net data rate takes into account only the user data. A measure of the efficiency of a bus system is the ratio of net data rate to gross data rate.

Access methods
A central function of a network to which several users have access at the same time is the administration of the token in order to avoid conflicts. This is known in buses as arbitration. It should be borne in mind here how urgently a message has to be transported. Different mechanisms are circulated.

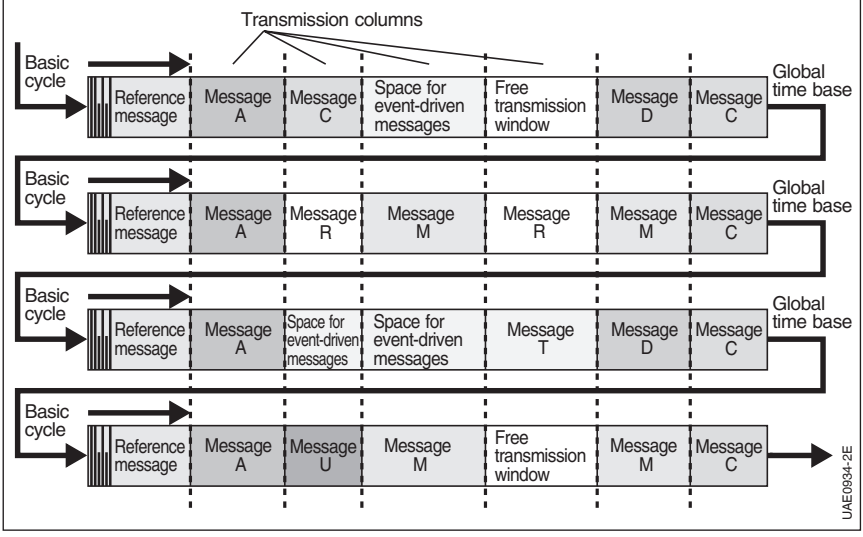
Time control
With time-controlled buses, conflict-free access is possible whereby each node receives permanently allocated times (slots, columns) for its emissions. These recurring times are combined into cyclesand the cycles combined into the communication matrix (Figure 4). This mechanism is called TDMA (Time Division Multiple Access).

The latency is limited in the case of time control to a maximum length. The response behavior of the bus system must therefore be predicted with regard to time.

The messages (reference messages, Sync Frame) themselves are used to synchronize the clocks in the nodes. If there are numerous timing nodes (time master) in the network, the clocks are synchronized on the basis of the measured reception times with the expected reception times. The strict periodicity (determinism) makes it possible to quickly detect when a message fails to materialize, which is required particularly in safety-relevant systems such as electric-motor-operated steering.

In addition, the time-controlled approach promotes the possibility that several subsystems are developed independently

Figure 4: Communication matrix of a time-controlled bus (example: TTCAN)

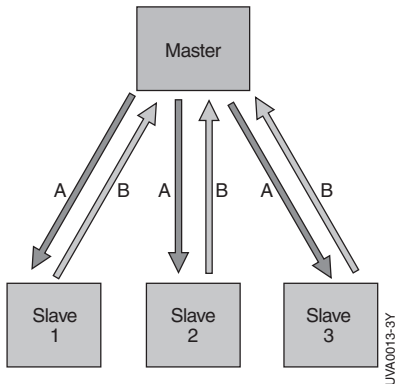


of each other, because in this way the composabilityof the subsystems into predictable behavior of the overall system is supported. When each subsystem has its independent time slice, two nodes cannot disrupt each other by simultaneously occupying the bus.

Master-slave
A designated node (master) grants one of the other nodes (slave) limited access (Figure 5). In this way, it determines the communication frequency by interrogating its subordinate nodes. A slave responds only when it is addressed by the master. This method produces defined latencies.

However, some master-slave protocols permit a slave to register with the master in order to send a message.

Figure 5: Master-slave method
A Master inquiry,
B Slave response.



Multi-master
In a multi-master network, various nodes can access the bus automatically and send a message if the bus appears free. In other words, each node can perform the master role and all the nodes can equally start a message transfer. However, this also means that methods for identifying and dealing with access conflicts must be in place. This can be done, for example, by way of a decision phase with prioritization or a delayed resending.

The use of priority control prevents a bus conflict when several nodes want to occupy the bus at the same time. The network node which has a high priority or which would like to transmit a high-priority message asserts itself and sends its message first. When the line is free again, all the nodes which have to send messages – particularly the node that is waiting – start with a renewed attempt. Consequently a maximum value cannot be specified for the latency with this method.

The multi-master architecture has a positive effect on system availability because communication is not controlled by any single node, the failure of which would result in a total communication failure.

Token passing
The right of access is temporarily awarded to a node, which then passes it on to another node for a limited time. The controlled allocation of the right of access ensures that only one single node has right of access to the bus at a given time.

Forms of communication
The simplest form of communication is the 1:1 connection (point-to-point or unicast connection), in which one transmitter and one receiver are connected via a bus.

When a transmitter sends a telegram to several receivers without all the receivers processing this telegram, this is called a broadcast connection. In a multicast connection each receiver processes the telegram provided by the transmitter.

Addressing
In order for messages to be transferred via a network and their information to be evaluated, they contain, in addition to the user data (payload), information on the data transfer. This information can be contained explicitly in the transfer or stipulated implicitly. Addressing is required to ensure that a message arrives at the correct receiver. There are different methods of doing this.

User-oriented method
Data are exchanged here on the basis of node addresses (Figure 6a). The message transmitted by the sender contains, as well as the data to be transferred, the address of the destination node. All the receivers compare the transmitted receiver address with their own and only those receivers with the correct address evaluate the message. Messages can be directed to individual addressees, to groups (multicast) or to all the nodes in the network (broadcast).

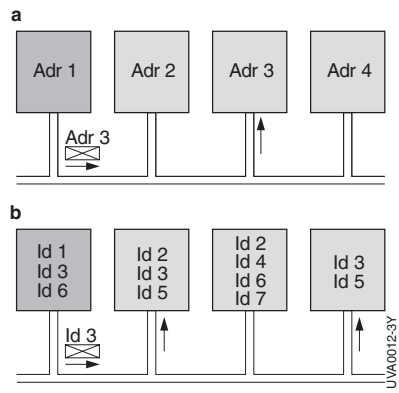
Most conventional communication systems (e.g., Ethernet) operate in accordance with the principle of User addressing

Message-oriented method
With this method, it is not the receiver nodes but the messages themselves which are addressed (Figure 6b). A message is identified according to its content by a message identifier which was established in advance for this information type. With this method, the sender does not require any knowledge of the destination of the message, since each receiver node decides for itself whether to process the message. Several nodes can also accept and evaluate the message.

Most bus systems in the automotive field (e.g., CAN, FlexRay) operated according to the message-oriented method.

Transfer-oriented method
Even transfer features can be used to identify a message. If a message is always sent in a defined time slot, it can be identified using this item. For safeguarding purposes, this addressing is also combined in many cases with a message- or user-oriented addressing.

Figure 6: Addressing types
a) User-oriented method,
b) Message-oriented method.
Adr i, $i = 1…n$, address of nodes.
Id i, $i = 1…n$, message identifiers.



Network topology
A network topology refers to the structure of network nodes and connections. This shows which nodes are connected to each other, but not the details, such as the length of the connection for example. For different applications of communication networks, there are different requirements of the topology used.

Bus topology
This network topology is also referred to as a linear bus. The core element is a single cable, to which all the nodes are connected by way of short connecting cables (Figure 7a). In this topology, it is very easy to expand the network by additional users. Messages are sent out from the individual bus users and distributed to the entire bus.

If a node fails, the data expected from this node are not available to the other nodes in the network, but the remaining nodes can continue to exchange messages. However, a network with a bus topology will fail completely if the central cable has a fault (e.g., in the event of a cable break).

Star topology

The star topology consists of a central node (repeater, hub, star), to which all the other nodes are connected by way of individual connections (Figure 7b). A network with this topology can therefore be easily expanded if free connections are available at the central element.

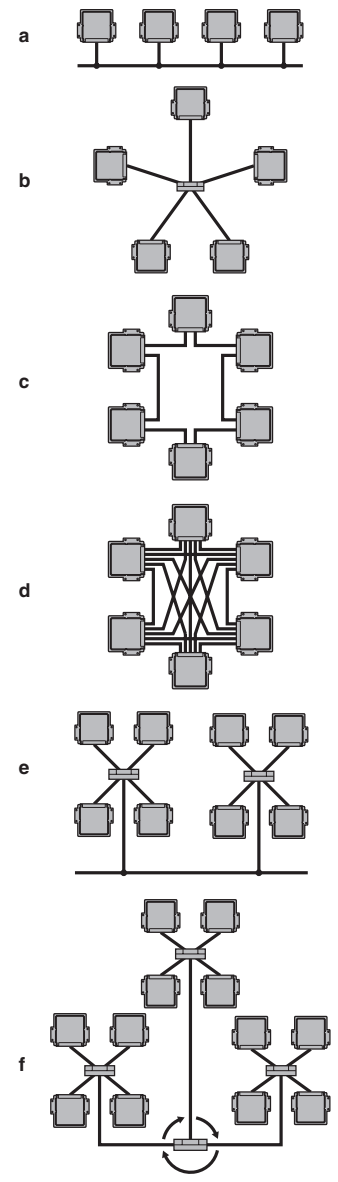
Data are exchanged via the individual node connections with the central star, where a distinction is made between active and passive stars. The active star contains a computer which processes and forwards the data. The network's performance is significantly determined by the performance of this computer. The central node, however, does not necessarily have to have special control intelligence. A passive star only joins the bus lines of the network users.

If one node fails or a connecting cable to the central node is faulty, the remaining network continues to be operational. If, on the other hand, the central node fails, then the entire network is rendered inoperational.

Ring topology

In the ring topology, each node is connected to its two neighbors. This creates a closed ring (Figure 7c). In a ring, data are transferred in only one direction from one station to the next. The data are checked in each case after they are received. If they are not intended for this station, they are renewed (repeater function), amplified, and sent forward to the next station. The data to be transferred are forwarded in the ring, i.e., from one station to the next, until they reach their intended destination or arrive at the starting point again. When a message has passed through the complete ring, reception is acknowledged by all the nodes. Once one station in a single ring fails, data transfer is interrupted and the network fails completely.

Figure 7: Network topologies
a) Bus topology, b) Star topology,
c) Ring topology, d) Mesh topology,
e) Star-bus topology,
f) Star-ring topology.



Rings can also be structured in the shape of a double ring in which data are transferred in both directions. In this topology, the failure of one station or of a connection between two stations can be managed, since all the data continue to be transferred to all the operational stations in the ring.

Daisy-chain topology
The daisy-chain topology looks like the ring topology when a connection is removed. Here, the first component is connected directly to data-processing equipment (e.g., a computer). The subsequent components are connected to their preceding components (series-connection principle), thereby creating a chain. Messages thus pass through several nodes until the destination is reached.

Mesh topology
In a mesh topology, each node is connected to one or more further nodes (Figure 7d). In the event of the failure of a node or a connection, there are diversions along which the data can be routed. This network is therefore characterized by high failure tolerance. However, the costs of networking and transporting the messages are high.

Hybrid topologies
In hybrid topologies, different network topologies are connected. The following combinations are possible, for example:
– Star-bus topology: The hubs of various star networks are connected to each other as a linear bus (Figure 7e).
– Star-ring topology: The hubs of various star networks are connected to the main hub (Figure 7f). In this main hub, the hubs of the star networks are connected in a ring shape.

Coupling of bus systems
Depending on the requirement, different bus systems with different communication protocols are used in motor vehicles. Intermediate stations are required to be able to transfer messages from one bus system to the other.

Repeater
Similar buses are coupled via the repeater. The repeater compensates the attenuation on the lines; it therefore works as a booster. The communication traffic is transferred from one bus system to the other.

Gateway
The gateway processes the incoming data packets and generates from them the data packets for the bus system to be operated. The gateway thus converts the incoming data in order to send it with the required communication protocol.

Router
The router forwards to the other bus only those data which are needed by the ECUs connected to this bus (nodes).

Standby state
With a command the node can be switched to a power-saving standby state, the sleep mode, when no data exchange is required. Communication is deactivated in this state.

The node is returned to the normal operating state with a wake-up. The different bus systems apply their specific methods to control the sleep mode and the wake-up mode.

Transfer reliability
An important feature of buses, particularly if they are to be used in safety-relevant systems, is their ability to identify faults and, if necessary, to maintain limited operation. This includes the possibility of detecting the corruption of data. Electromagnetic radiation from the surroundings (e.g., from the ignition coil) on the cables may cause some bits to assume incorrect values at the receiver. These incidents can be identified by transferring check information in addition to the data.

The simplest case is an additional parity bit, which is used to detect whether the correct number of ones is even or uneven. A more elaborate method is the block-securing method (cyclic redundancy check, CRC), in which different levels of protection can be achieved with a selectable number of check bits. The transmitting node generates via the encoding rule the check information and thereby supplements the telegram to be sent. The receiver calculates from the received telegram the check information according to the same algorithm and compares it with the received value. A data-transfer error can be detected in this way.

One way of ensuring limited operation is to bypass failed lines and prevent blockage of the bus by faulty nodes.

It is ensured that an item of data sent out by a node, e.g., the vehicle speed, is available either to all the addressees in the network or to none (consistency). The situation where two nodes have different images of the speed value therefore cannot arise. For this purpose, a node signals that reception has failed or refrains from acknowledging correct reception (acknowledge). The other nodes then reject the data already correctly received by them.

Buses in motor vehicles

External buses connect ECUs, sensors, and actuators to each other. They run via a wiring harness and plug-in connections. Levels, bit rates, and impedances are precisely defined in order to ensure interoperability. Examples are CAN, FlexRay, and LIN. For a quick overview, these buses are frequently classified according to the scheme in Table 1.

Internal buses connect integrated circuits on the board. They are only used inside an ECU.

Table 1: Classification of bus systems

Class A	
Data transfer rates	Low data rates (up to 10 kbit/s)
Applications	Networking of actuators and sensors
Represented by	LIN, PSI5
Class B	
Data transfer rates	Medium data rates (up to 125 kbit/s)
Applications	Complex mechanisms for error handling Networking of control units in the comfort area
Represented by	Lowspeed-CAN (CAN-B)
Class C	
Data transfer rates	High data rates (up to 1 Mbit/s)
Applications	Real-time requirements Networking of control units in the drive and chassis area
Represented by	High-speed CAN (CAN-C)
Class C+	
Data transfer rates	Very high data rates (up to 10 Mbit/s)
Applications	Real-time requirements Networking of control units in the drive and chassis area
Represented by	FlexRay
Class D	
Data transfer rates	Very high data rates (from 10 Mbit/s)
Applications	Networking of control units in the areas of telematics and multimedia
Represented by	MOST, Ethernet

CAN

Overview

The CAN bus (Controller Area Network) has established itself as the standard since its first series introduction in motor vehicles in 1991. But it is also often used in automation technology. The major features are:

- Priority-controlled message transmission with non-destructive arbitration.
- Low costs through the use of a low-cost twisted two-wire cable and use of a simple protocol with low computing-power demand.
- A data-transfer rate up to 1 Mbit/s for the high-speed CAN (CAN-C) and up to 125 kbit/s for the low-speed CAN (CAN-B, lower expenditure for the hardware).
- High reliability of data transfer through recognition and signaling of sporadic faults and permanent faults and through network-wide consistency via Acknowledge.
- The multi-master principle.
- High availability by locating failed stations.
- Standardization in accordance with ISO 11898 [1].

Transfer system

Logic bus states and encoding

To communicate, the CAN bus uses the two states “dominant” and “recessive”, with which the information bits are transmitted. The dominant state represents “0”, the recessive “1”. The NRZ process (Non Return to Zero) is used as encoding for transmission, in which a zero state is not always returned to between two equal transfer states and, therefore, the time interval between two edges which is required for synchronization can become too great.

For the most part a two-wire cable with, depending on the ambient conditions, a non-twisted or twisted pair is used. The two bus lines are called CAN_H and CAN_L (Figure 1).

The two-wire cable facilitates a symmetrical data transfer, in which the bits are transferred via both bus lines using different voltages. This reduces sensitivity to common-mode interference, since interference affects both lines and can be filtered out by creating the difference (Figure 2).

The single-wire cable represents a way of lowering manufacturing costs by saving on the second cable. However, a common ground connection which assumes the function of the second cable must be available to all the bus users for this purpose. The single-wire version of the CAN bus is therefore only possible for a communication system with limited spatial reach. Data transfer on the single-wire cable is more susceptible to interference radiation, since it is not possible to filter out interference pulses as on the two-wire cable. A higher level signal on the bus line is therefore required. This in turn has a negative effect on the interference radiation. The edge steepness of the bus signals must therefore be reduced compared with the two-wire cable. This is associated with a lower data-transfer rate. For this reason, the single-wire cable is only used for the low-speed CAN in the area of body and comfort and convenience electronics. The single-wire solution is not described in the CAN specification.

There are also so-called fault tolerant transceivers, which communicate via a two-wire cable, but if a cable breaks continue operating as a single-wire system.

Voltage levels

High-speed and low-speed CANs use different voltage levels to transfer dominant and recessive states. The voltage levels of the low-speed CAN are shown in Figure 1a, those of the high-speed CAN in Figure 1b.

The high-speed CAN uses, in the recessive state on both lines, a nominal voltage of 2.5 V. In the dominant state, nominal voltages of 3.5 V and 1.5 V are applied to CAN_H and CAN_L respectively. In the low-speed CAN in the recessive state, a voltage of 0 V (max. 0.3 V) is applied to CAN_H and a voltage of 5 V (min. 4.7 V) is applied to CAN_L. In the dominant state, the voltage at CAN_H is min. 3.6 V and at CAN_L max. 1.4 V.

Figure 1: Voltage level of CAN data transfer

a) Low-speed CAN (CAN-B),
b) High-speed-CAN (CAN-C).
CAN_H CAN High level,
CAN_L CAN Low level.

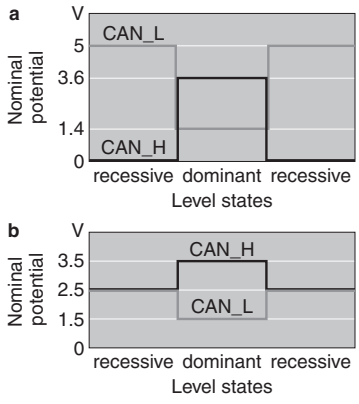
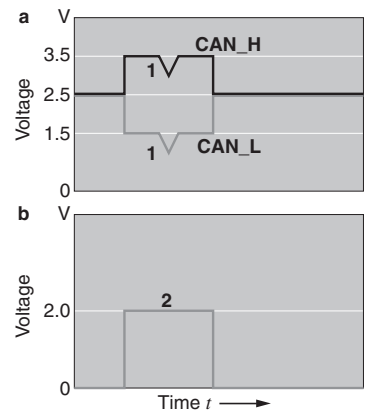


Figure 2: Filtering out of interference on the CAN bus

a) Signal level of CAN lines with interference on both lines,
b) Difference signal.
1 Interference pulse,
2 Difference signal.
CAN_H CAN High level,
CAN_L CAN Low level.



Limits
It is important for the arbitration method in the case of CAN that all the nodes in the network see the bits of the message identification (frame identifier) simultaneously so that a node – while it is still transmitting a bit – sees whether another node is also transmitting. Delays result from the signal propagation time on the data bus and the processing times in the transceiver. The maximum permissible transfer rate is thus dependent on the overall length of the bus. ISO specifies 1 Mbit/s for 40 m. For longer lines, the possible transfer rate is roughly inversely proportional to the line length. Networks with a 1 km reach can be operated with 40 kbit/s.

CAN protocol
Bus configuration
CAN operates according to the multimaster principle, in which a linear bus topology connects several nodes of equal priority rating.

Content-based addressing
CAN uses message-based addressing. This involves assigning a fixed identifier to each message. The identifier (ID) classifies the content of the message (e.g. engine speed). Each station processes only messages whose identifiers are stored in the list of messages to be accepted. This is called acceptance checking (Figure 3). Thus, the CAN requires no station addresses for data transmission, and the nodes are not involved in managing the system configuration. This facilitates adaptation to variations in equipment levels.

Logic bus states
The CAN protocol is based on two logic states: The bits are either “recessive” (logic 1) or “dominant” (logic 0). When at least one station transmits a dominant bit, it overwrites the recessive bits simultaneously sent by other stations.

Bus arbitration and priority assignments
Each station can begin transmitting a message as soon as the bus is unoccupied. When several stations start to transmit simultaneously, a “wired-and arbitration scheme” becomes active to resolve the resulting bus-access conflicts. The arbitration scheme ensures that the domi-

Figure 3: Addressing and acceptance checking
Station 2 sends, stations 1 and 4 receive the data.

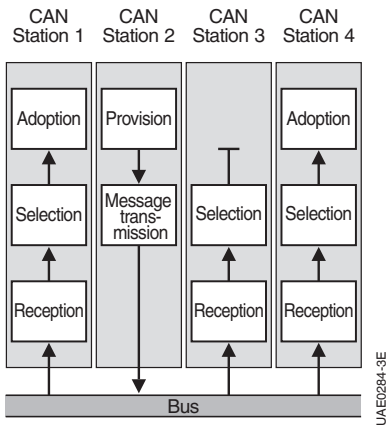
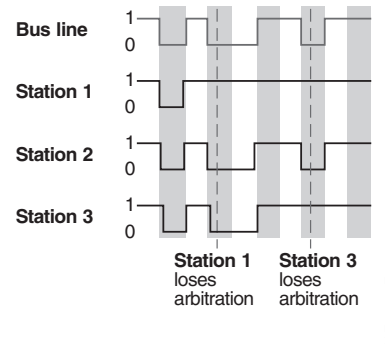


Figure 4: Bitwise arbitration
0 Dominant level,
1 Recessive level.
Station 2 gains first access (signal on the bus corresponds to signal from station 2).

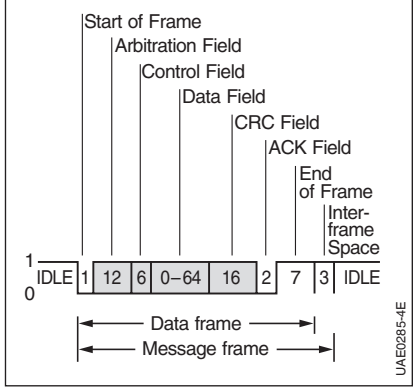


nant bits transmitted by a station overwrite the recessive bits of other stations (Figure 4). Each station gives bit by bit – the highest significant bit first – the identifier of its message on the bus. During this arbitration phase (selection phase) each transmitting station compares the applied bus level with the actual level. Each station that transmits a recessive bit but observes a dominant bit loses arbitration. The station with the lowest identifier – i.e. with the highest priority – is assigned first access on the bus without having to repeat the message (non-destructive access control). Transmitters respond to failure to gain bus access by automatically switching to receive mode; they then repeat the transmission attempt as soon as the bus is free again.

Data frame and message format

The CAN supports two different message formats, which primarily differ with respect to the length of their identifiers. The standard format has 11 bits, while the extended format has 29 bits. Thus, the data frame to be transmitted contains a maximum of 130 bits (in the standard format) and 150 bits (in the extended format). This ensures minimum waiting time until the next transmission (which could be urgent). The data frame for the standard format consists of seven consecutive fields (Figure 5). The "Start of Frame" indicates the beginning of a message and synchronizes all nodes.

Figure 5: Data frame



The "Arbitration Field" consists of the message's identifier and an additional control bit. While this field is being transmitted, the transmitter accompanies the transmission of each bit with a check to ensure that no higher-priority message is being transmitted (which would cancel the access authorization). The control bit determines whether the message is classified under "data frame" or "remote frame".

The "Control Field" contains the code indicating the number of data bytes in the "Data Field".

The "Data Field" has an information content of between 0 and 8 bytes. A message of data length 0 can be used to synchronize distributed processes.

The "CRC Field" (Cyclic Redundancy Check) contains a checksum for detecting possible transmission interference.

The "Ack Field" contains the acknowledgement signals with which the receivers indicate the error-free receipt of messages.

The "End of Frame" marks the end of the message.

Then comes the "Interframe Space" to provide a separation from the next data frame.

Transmitter initiative

The transmitter will usually initiate a data transmission by sending a data frame. However, the receiver can also request data from the transmitter. This involves the receiver sending off a "Remote Frame". The "Data Frame" and the corresponding "Remote Frame" have the same identifier. They are distinguished by the bit that follows the identifier.

Error detection
The CAN incorporates a number of monitoring features for detecting errors. These include:
– 15-bit CRC: Each receiver compares the CRC sequence which it receives with the calculated sequence.
– Monitoring: Each transmitter reads its own transmitted message from the bus and compares each transmitted and scanned bit.
– Bit stuffing: Between “Start of Frame” and the end of the “CRC Field”, each “Data Frame” or “Remote Frame” may contain a maximum of five consecutive bits of the same polarity. The transmitter follows up a sequence of five bits of the same polarity by inserting a bit of the opposite polarity in the bitstream. The receivers eliminate these bits again as the messages arrive.
– Frame check: The CAN protocol contains several bit fields with a fixed format for verification by all stations.

Error handling
When a CAN controller detects an error, it aborts the current transmission by sending an error flag. An error flag consists of six dominant bits; it functions by deliberately violating the stuffing convention and the formats.

Fault confinement with local failure
Defective stations can severely impair the ability to process bus traffic. Therefore, the CAN controllers incorporate mechanisms which can distinguish between intermittent and permanent errors, and local station failures. This process is based on statistical evaluation of error conditions.

Implementations
The semiconductor manufacturers offer different implementations of CAN controllers which differ primarily in the extent to which they can store and manage messages. In this way, the host computer can be relieved of protocol-specific activities. A standard categorization is Basic CAN controllers, which have only a few message memories, and Full CAN controllers, in which there is room for all the messages necessary for an ECU.

Standardization
The CAN has been standardized for data exchange in motor vehicles; for applications with a low transfer rate up to 125 kbit/s as ISO 11898-3 [1] and for applications with a high transfer rate over 125 kbit/s as ISO 11898-2 [1] and SAE J 1939 (truck and bus, [2]).

Time-triggered CAN
The extension of the CAN protocol to include the capability of operating in time-triggered mode is called “Time-Triggered CAN” (TTCAN). It is fully configurable with regard to the proportion of time-triggered to event-driven communication components, and is therefore fully compatible with CAN networks. TTCAN is standardized as ISO 11898-4 [1].

CAN with flexible data rate
CAN-FD extends CAN by a second bit rate and a wider data field. Unlike the nominal bit rate previously applicable to CAN, the data bit rate is not limited to 1 Mbit/s. The data bit rate acts exclusively on the data field within the data frame, whereas the nominal bit rate still acts on the control data.

In addition, CAN-FD increases the data field from 8 to up to 64 bytes. This also necessitates adaptations to the checksum so as not to compromise transmission reliability. CAN-FD is standardized as ISO 11898-7 [1].

FlexRay

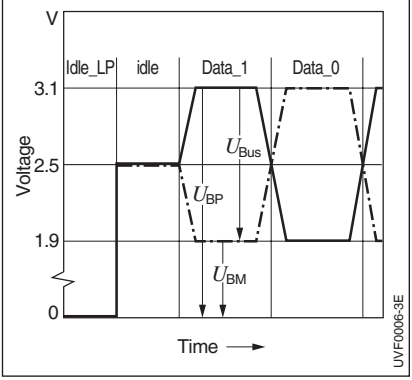
Overview

FlexRay is a bus which was designed for control engineering in the automotive field. Particular attention was given in development to its suitability for use in active safety systems without a mechanical fallback level (X-by-Wire), where determinism and fault tolerance are required. Essentially, thanks to the high transfer rate of up to 20 Mbit/s for non-redundant transmissions, its use in the field of audio transmission or for highly compressed video transmission is also conceivable. The major features are:

– Time-triggered transmission with guaranteed latency,
– Possibility of event-driven transmission of information with prioritization,
– Transmission of information via one or two channels,
– High transfer rate of up to 10 Mbit/s, with parallel transfer via two channels up to 20 Mbit/s,
– Structure as a linear bus, in a star configuration or as a mixed form.

FlexRay is the first automotive communication standard to have been created in a consortium of vehicle manufacturers, suppliers and semiconductor manufacturers. It contains elements from TTCAN, Byteflight and other technologies. The specifications published by the FlexRay Consortium are today (only) available as ISO standard 17458 [3].

Figure 6: Bus states and voltages for FlexRay



Transmission media

The transmission medium used in a FlexRay system is a twisted-pair two-wire cable where both shielded and non-shielded cables can be used. Each FlexRay channel consists of two strands, Bus-Plus (BP) and Bus-Minus (BM). FlexRay uses NRZ (Non Return to Zero) for encoding.

The bus state is identified by measuring the voltage difference between Bus-Plus and Bus-Minus. Data transfer is thus less sensitive to external electromagnetic influences, since these act equally on both strands and cancel each other out in the difference.

When different voltages are applied to the two strands of a channel, four bus states can be assumed; these bus states are referred to as Idle_LP (LP, Low Power), Idle, Data_0, and Data_1 (Figure 6). Idle_LP is the state in which a low voltage between –200 mV and 200 mV (to ground) is applied to Bus-Plus and Bus-Minus. In the Idle state, a voltage of 2.5 V with a maximum difference of 30 mV is applied to BP and BM. In order to place the channel in the Data_0 state, at least one transmitting node must apply a negative differential voltage of –600 mV to the channel, for Data_1,600 mV, based on a mid-level of 2.5 mV.

Topologies

FlexRay networks can be structured as bus topologies and star topologies. Two stars can be cascaded – when signal delays are taken into consideration in the stars. Topologies in which several buses are connected to a star are also possible.

Because both channels of a FlexRay system can be structured independently of each other, it is possible for different topologies to be used for both channels. For example, one channel can be structured as an active star topology, the other as a bus topology.

On account of the frequencies, which can be ten times those of CAN, it is necessary when designing a FlexRay network to ensure in all topologies in particular that parameters such as line length and terminating resistors are selected in such a way that signal distortions remain in the permissible range.

Bus access, time control

In order to achieve determinism, i.e. the guarantee of a maximum duration for the transmission of a message, communication is performed in the FlexRay bus on a time-controlled basis in cycles of constant duration. Each cycle first features a static segment which is divided into time slots of equal length (Figure 7). Each time slot is permanently assigned a maximum of one node which is permitted to transmit at this time.

This is followed by a dynamic segment in which bus access is regulated by the priority of messages. The split between the static and dynamic segments is freely configurable, but cannot be altered during operation. The same applies to the lengths of the time slots, which are configurable, but must remain constant during operation.

The "Symbol Window" can optionally be defined as the third element in the cycle. This can be used to transmit a single symbol. Symbols are provided to wake a network and to test functionalities.

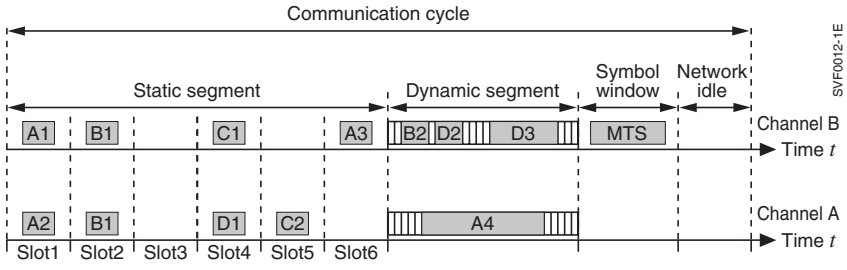
Synchronization

Each network node requires its own time generator, which it utilizes to decide the time for transmitting and to determine the duration of the bits. The internal time generators of several nodes may deviate from each other on account of temperature and voltage fluctuations and manufacturing tolerances. In a bus system such as FlexRay, which controls bus access via time slots, it is therefore necessary to ensure that the deviation of the clocks from one another remains within a permissible range through regular corrections. To this end, some nodes assume the role of time generators, to which the other nodes regularly synchronize their internal clocks. The procedure adapts both the zero points (offset) of the clocks and their rate. It can continue to operate even if individual nodes fail. To be able to make corrections, each cycle ends with a short phase (NIT, Network Idle Time) in which the zero point of the cycle can be shifted.

Thanks to this procedure a "global time" is provided in all the nodes; this time is given in macroticks. The synchronization mechanism causes the length of one macrotick to be the same on average in all the nodes.

When the network is switched on, a common conception of time on the part of all the nodes must first be established. The start-up process, which takes a little time, serves this purpose. Likewise, a node which intends to synchronize itself to a running network needs a period of time to be taken into consideration.

Figure 7: Time control for FlexRay (example)
A1 Node A transmits message 1,
A2 Node A transmits message 2,
MTS Media Test Symbol.

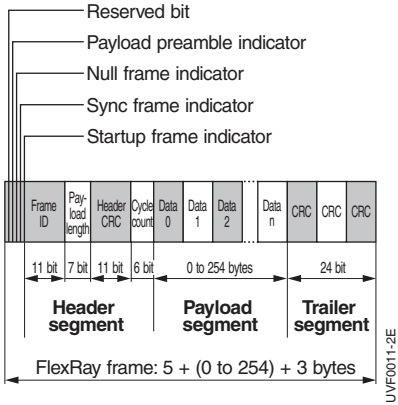


Arbitration in the dynamic segment
Messages can be given different priorities in the dynamic segment. The duration until a message is transmitted cannot be guaranteed for this. The priority is established by the frame ID, which may be allocated only once in the network. The messages are transmitted in the sequence of their frame ID. To this end, each node runs a counter (slot ID), which is increased when a message is received. If the slot ID assumes the value of the frame ID of a message ready in this node, it is transmitted. If the length of the dynamic segment is not sufficient for all the messages, the transmission process must be shifted to a later cycle.

The data frames in the dynamic segment can have different lengths. The limits of the dynamic slots on the two channels are independent of each other. Thus, there may be messages with different slot IDs on the channels at one time.

Data frame
FlexRay uses in both the static and dynamic segments the same data-frame format, which can be divided into the three sections of Header, Payload and Trailer (Figure 8).

Figure 8: Data frame



Header
The header comprises:
– The reserved bit for future protocol changes.
– The payload preamble indicator, which identifies whether the payload contains a network management vector.
– The null frame indicator, which identifies that the data have not been updated since the last cycle.
– The sync frame indicator, which signals that this data frame is to be used to synchronize the system.
– The startup frame indicator, which identifies that this frame is used in the network's start-up phase.
– The frame ID; this corresponds to the number of the slot in which the data frame is transmitted.
– The payload length which contains the size of the user data. For all the slots in the static segment this field always contains the same value. Data frames in the dynamic segment can have different lengths.
– The header CRC, which provides this part of the data frame with extra protection, because it is crucial to the time response.
– The cycle count; the number of the cycle in which the transmitting network node is located is transmitted in this field.

Payload
The user data which are processed further by the host are transmitted in the payload segment. For data frames in the static segment the first payload bytes can optionally be declared to the network management vector. The controllers or all the vectors received in the cycle and make them accessible to the host. For data frames in the dynamic segment the first payload bytes can optionally be declared to a 16-bit message ID. In both cases, subsequent handling is left to the software.

The user data have a maximum length of 254 bytes, which are transmitted in 2-byte words.

Trailer
The trailer contains a 24-bit checksum (frame CRC), which acts on the entire data frame.

Generating a frame bitstream
Before a node can transmit a data frame with the host's data, the data frame is converted into a "bitstream". To this end, the data frame is first deconstructed into individual bytes. A Transmission Start Sequence (TSS) of configurable bit length is inserted at the start of the data frame, followed by a bit Frame Start Sequence (FSS). Then an extended byte sequence is generated from the bytes of the data frame, where a two bit Byte Start Sequence (BSS) is inserted ahead of each frame byte.

To conclude the bitstream, a two bit Frame End Sequence (FES) is appended to the bitstream.

In the event that the data frame is in the dynamic segment, it is possible to append to the bitstream a further Dynamic Trailing Sequence (DTS) of configurable bit length. This prevents another node from starting its transmission via the channel early.

Operating modes
FlexRay can be placed in a mode in which the nodes need only minimal power and in which all the operations of the encoding and decoding process are stopped, but can be woken up by a signal on the bus line. Here, the bus driver is still capable of detecting special signals on the bus and then also of activating its host by means of a corresponding signal. Each node can transmit a wake-up.

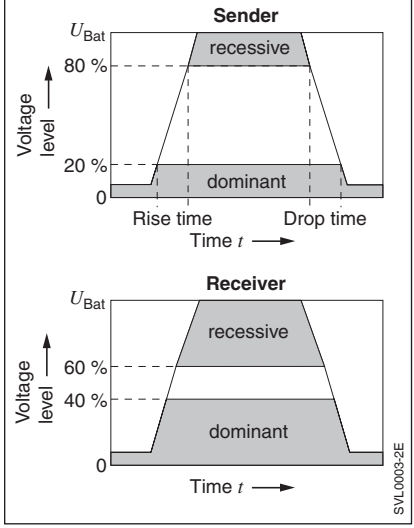
LIN

Overview
The LIN bus (Local Interconnect Network) was designed to cover the communication requirements for Class A systems (see Table 1) with the most cost-effective hardware possible in the node. Typical applications are the door module with door lock, the power-window units, door-mirror adjustment, and air-conditioning system (transmission of signals from the control element, activation of the fresh-air blower).

The current LIN specification can be accessed on the LIN Consortium website [4].

Important features of the LIN bus are:
– Single-master and multi-slave concepts,
– Low hardware costs on account of data transfer via non-shielded single-wire cable,
– Self-synchronization of the slaves even without a quartz oscillator,
– Communication in the form of very short messages,
– Transfer rate max. 20 kbit/s,
– Bus length up to 40 m and up to 16 nodes.

Figure 9: Voltage level on data line of LIN bus



Transfer system

The LIN bus is designed as a non-shielded single-wire cable. The bus level can assume two logic states. The dominant level corresponds to a voltage of approximately 0 V (ground) and represents logic 0. The recessive level corresponds to the battery voltage U_{batt} and represents the logic 1 state.

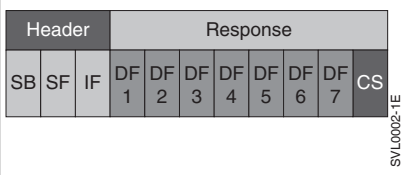
Because of the different design variations of the circuitry, the levels may be subject to differences. The definition of tolerances for transmitting and receiving in the field of the recessive and dominant levels ensures a stable data transfer. The tolerance bands are wider at the receiving end (Figure 9) so that valid signals can also be received in spite of interference radiation.

The transfer rate of the LIN bus is limited to 20 kbit/s. This is a compromise between the demand for high edge steepness to synchronize the slaves easily on the one hand and the demand for low edge steepness to improve EMC performance on the other. Recommended transfer rates are 2,400 bit/s, 9,600 bit/s and 19,200 Bit/s. The minimum permissible value for the transfer rate is 1 kbit/s.

The maximum number of nodes is not stipulated in the LIN specification. Theoretically, it is limited by the number of available content-related message identifiers. Line and node capacities and edge steepnesses limit the combination of length and node number of a LIN network; a maximum of 16 nodes is recommended.

The bus users are usually arranged in a linear bus topology; however, this topology is not explicitly prescribed.

Figure 10: LIN frame
SB Synch Break,
SF Synch Field,
IF Ident Field,
DF Data Field,
CS Checksum.



Bus access

In the LIN bus, access is provided on the basis of the master-slave access method. The network features a master, which initiates each message. The slave has the opportunity to respond. The messages are exchanged between the master and one, several or all the slaves.

The following relationships are possible during communication between master and slave:

– Message with slave response: The master transmits a message to one or more slaves and asks for data (e.g. switch states of measured values).
– Message with master instruction: The master issues control instructions to a slave (e.g. switching on a servomotor).
– Message for initialization: The master initiates a communication between two slaves.

LIN protocol

Data frame

The information on the LIN bus is embedded in a defined data frame (LIN frame) (Figure 10). A message initiated by the master begins with a header. The message field (response) contains different information depending on the type of message. If the master transmits control instructions for a slave, it describes the message field with the data to be utilized by the slave. In the event of a data request, the addressed slave describes the message field with the data requested by the master.

Header

The header is made up of the synchronization break (Synch Break), the synchronization field (Synch Field), and the identifier field (Ident Field).

Synchronization
A synchronization takes place at the start of each data frame to ensure a consistent data transfer between master and slaves. First the start of a data frame is clearly identified by the Synch Break. It consists of at least 13 consecutive dominant levels and one recessive level.

After the Synch Break, the master transmits the Synch Field, consisting of the bit sequence 01010101. The slaves thus have the opportunity to adapt themselves to the master's time base. The clock pulse of the master should not deviate from the nominal value by more than ±0.5%. The clock pulse of the slaves may deviate prior to synchronization by up to ±15% if the synchronization achieves a deviation of max. ±2% up to the end of the message. The slaves can thus be designed without an expensive quartz oscillator, for example with a cost-effective RC circuit.

Identifier
The third byte in the header is used as the LIN identifier. Similarly to the CAN bus, content-based addressing is used – the identifier therefore gives information about the content of a message. All the nodes connected to the bus decide on the basis of this information whether they intend to receive and process the message or ignore it (acceptance filtering).

Six of the eight bits of the identifier field determine the identifier itself, from which 64 possible identifiers (ID) are obtained. They have the following meanings:
– ID = 0–59: Transmission of signals,
– ID = 60: Master request for commands and diagnostics,
– ID = 61: Slave response to ID 60,
– ID = 62: Reserved for manufacturer-specific communication,
– ID = 63: Reserved for future expansions of the protocol.

Of the 64 possible messages, 32 may contain only two data bytes, 16 four data bytes, and the remaining 16 eight data bytes.

The last two bits in the Ident Field contain two checksums, with which the identifier is protected against transmission errors and resulting incorrect message allocations.

Data field
Transmission of the actual data begins after the master node has transmitted the header. The slaves identify from the transmitted identifier whether they are addressed and, if necessary, transmit back the response in the data field.

Several signals can be packed into a data frame. Here, each signal has exactly one generator, i.e. it is always described by the same network node. During operation it is not permitted to change the signal allocation to another generator, as would be possible in other time-controlled networks.

The data in the slave response are safeguarded by a checksum (CS).

LIN description file
The configuration of the LIN bus, i.e. the specification of network users, signals and data frames, is performed in the LIN description file. The LIN specification provides for a suitable configuration language for this purpose.

From the LIN description file, tools automatically generate program sections which are used to implement the master and slave functions in the ECUs located on the bus. The LIN description file thus serves to configure the entire LIN network. It is a common interface between the vehicle manufacturer and the suppliers of the master and slave modules.

Message scheduling
The scheduling table in the LIN description file determines the order and time frame in which the messages are transmitted. Frequently needed information is transmitted from time to time. When the table has been worked through, the master begins again with the first message. The sequence of processing can be altered depending on the operating state (e.g. diagnostics active or inactive, ignition on or off).

Thus, the transmission frame of each message is known. The deterministic performance is guaranteed by the fact that all the transmissions are initiated by the master in the case of master-slave access control.

Network management
The nodes of a LIN network can be placed in sleep mode in order to minimize closed-circuit current. Sleep mode can be achieved in two ways. The master transmits the "Go to Sleep" command with the reserved identifier 60, or the slaves automatically go into sleep mode if there has been no data transfer on the bus for an extended period of time (four seconds). Both the master and the slaves can wake up the network again. The wake-up signal must be transmitted for this purpose. This consists of a data byte with the number 128 as content. After a break of 4 to 64 bit times (wake-up delimiter), all the nodes must be initialized and be able to respond to the master.

Ethernet

Overview
The term Ethernet refers to a family of buses in which the addressing, the format of the messages and access control are identical (laid down in IEEE 802, [5]). Ethernet and the Internet Protocol (IP) were developed for data communication between computers or peripherals which are locally separate and where, during operation, reconfigurations of the network through the addition of new users or the failure of users can occur. The Ethernet buses are identified by the following important features:
– The transfer rate is in the range of 10 Mbit/s up to 10 Gbit/s.
– Data transfer is possible via assorted media such as coaxial cable, twisted two-wire cable, radio or glass fiber.
– The technology involved is standardized and very widely used.
– Simple insertion and removal of nodes is possible.
– The time response in the case of real-time applications is not guaranteed.

Ethernet is used in series-production vehicles; e.g. in the BMW 7 Series, where it is used to input vehicle data at the end of production.

Transfer system
The Ethernet versions differ in terms of transfer rate, physical design of the channel, and encoding. Coaxial cables, twisted two-wire cables with one or more core pairs, optical fibers, radio paths, or even power-supply cables are specified as channels. Encoding differs accordingly.

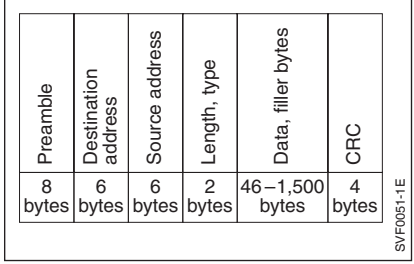
Originally, coaxial cables in a bus topology were used as the medium. Here, the transceivers of the nodes were connected either directly or with T-pieces to the cable. Today, twisted two-wire cables are widely used. The transfer rates have been increased from initially 10 Mbit/s via the Fast Ethernet with 100 Mbit/s and the Gigabit Ethernet with 1,000 Mbit/s up to 10 Gbit/s.

Topology

The size of a network is limited by the fact that the signal propagation time between two nodes influences the arbitration process. This can be bypassed by subdividing into segments, which are connected by way of special components – hubs and switches. A hub functions as an amplifier which re-establishes the ideal signal shape of a bit if it has been corrupted by interference or dispersion on the transmission medium. A switch checks entire packets for correctness with regard to the checksum and directs packets without collision to another output if the destination address can be reached by this route. To this end, it must have the option of storing messages temporarily. As well as the costs involved in the hardware, a disadvantage when such elements are used is that the data stream is delayed. For this purpose, however, nodes with different data rates can be connected.

Today, networks are usually designed in such a way that each node is connected to the output of a switch, i.e. there is no direct connection between nodes. Switches themselves are in turn connected via a higher-level switch in such a way that a tree-shaped structure is created.

Figure 11: Frame format of Ethernet protocol



Ethernet protocol

Bus access

To transmit, a node checks whether there are signals on the bus. It starts to transmit when it deems the line to be clear. Because of the signal propagation time between two nodes, the situation may arise where two nodes deem the bus to be clear and start to transmit virtually simultaneously. The data frames transmitted in the process are destroyed. The nodes identify this, abort their transmission, and wait a certain amount of time – different for each node – until they begin with a new transmission attempt. This destruction of data frames reduces the effective transfer rate to a tolerable extent provided bus utilization is not too high.

This arbitration process limits the length of messages and the propagation time, i.e. the reach. There are no priorities among the messages. A maximum duration for transmission can, therefore, not be guaranteed.

Each node adopts from all the messages those which contain its own as the destination address for further processing.

Data frame

Figure 11 shows the slightly simplified structure of a data frame. The preamble is a periodic bit sequence (101010 – 1011) and thereby generates a signal to synchronize the receiver. Messages contain the address of their source and their destination. Each network card has a unique address. The receiving nodes compare the destination address with their own card address and accept the data frame if they match. With multicast and broadcast addresses, several receivers can also be addressed.

PSI5

Overview

The Peripheral Sensor Interface 5 (PSI5) is a digital interface published by the PSI Consortium ([6]) for sensor applications in motor vehicles and can be assigned by the application to Class A (see Table 1). The PSI5 is based on already existing interfaces for peripheral airbag sensors, but has been developed as an open standard which can be used and implemented at no extra charge. The technical characteristics mentioned in the following, the low implementation expenditure and the low additional costs compared with an analog sensor connection make the PSI5 an attractive option for sensor applications in motor vehicles too.

Transfer system

The PSI5 is a two-wire current interface in which the same lines are used to power the sensors and for Manchester-encoded data transfer. The bus master in the ECU modulates a voltage to the sensor for this purpose. Data transfers from the sensor to the ECU are effected by current modulation of the supply lines. In this way, high EMC robustness and low electromagnetic radiation are achieved. A wide range of supply currents for the sensors can be supported.

The different PSI5 operating modes define the topology and parameters of communication between ECU and sensors (Figure 12):

- Communication modes: The asynchronous mode can be used for a unidirectional point-to-point connection. In the three synchronous bus modes (parallel, universal or daisy-chain cabling) several sensors can communicate on a time-controlled basis with the bus master bidirectionally and using TDMA processes.
- Data-word width: The PSI5 supports a variable data-word width of 8, 10, 16, 20 or 24 bits.
- Error detection: This can be effected either with a parity bit with even parity or with three CRC checksum bits.
- Cycle times: These are specified in µs.
- Number of time slots per cycle.
- Data-transfer rate: As standard 125 kbit/s or optionally 189 kbit/s.

For example, the operating mode “PSI5-P10P-500/3L” denotes a parallel synchronous mode with ten bits per data word and parity bit for error detection. The data are transmitted every 500 µs with three time slots per cycle and a low bit rate.

Figure 12: Designation of PSI5 operating modes

A Asynchronous mode,
P Parallel, synchronous mode,
U Universal cabling,
D Daisy-chain cabling,
P Parity bit,
CRC Cyclic redundancy check,
L Low (low bit rate),
H High (high bit rate).

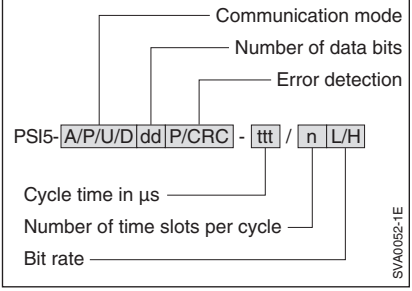
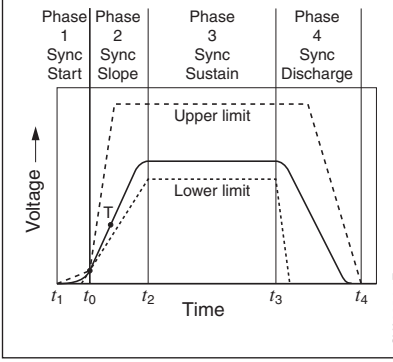


Figure 13: Synchronization signal of PSI5

T Trigger point.



During communication from sensor to ECU, a “low level” is represented by a normal (non-oscillating) current input of the sensor. A “high level” is generated by an increased current sink of the sensor (typically 26 mA). This current modulation is detected by the receiver in the ECU.

Each PSI5 data packet consists of N bits, in which there are in each case two start bits and one parity bit (or three CRC bits) and N–3 (or N–5) data bits. The data bits are transmitted with the least significant bit (LSB) first. Error detection with a parity bit is recommended for eight or ten bits, three CRC bits for longer data words.

In a PSI5 message, the data and value ranges have different meanings. One range serves to transmit the sensor output signals (≈ 94 %), one range the status and error messages (≈ 3 %), and one range the initialization data (≈ 3 %).

After each starting or undervoltage reset, the sensor conducts an internal initialization, after which the sensor is in an executable mode.

Whereas communication by the sensor with the ECU is effected with current signals, voltage modulation of the supply lines is used during communication by the ECU with the sensors. A logic 1 is represented by the synchronization signal, a logic 0 by the absence of the expected synchronization signal within the designated time slot. The synchronization signal consists of four voltage phases (Figure 13):

– Sync Start (nom. 3 µs, < 0.5 V),
– Sync Slope (nom. 7 µs, > 0.5 V),
– Sync Sustain (nom. 9 µs, > 3.5 V), and
– Sync Discharge (nom. 19 µs, < 3.5 V).

MOST

Overview

Application

The MOST bus (Media Oriented Systems Transport) was specially developed for the networking of multimedia applications in motor vehicles (infotainment bus). Alongside the classical entertainment functions such as radio receivers and CD players, infotainment systems also provide video functions (DVD and TV), navigation functions, and access to mobile communication and information. The MOST bus supports the logical networking of up to 64 devices and provides a fixed and reserved transmission bandwidth. MOST defines the protocol, the hardware, the software, and the system layers. MOST is jointly developed and standardized by automobile manufacturers and suppliers within the MOST Cooperation [7]. With a data rate of more than 10 Mbit/s, the MOST bus is a Class D bus system (see Table 1).

For data transfer, the MOST bus supports the following transmission channels:

- Control channel to transport control commands,
- Multimedia channel (synchronous channel) for the transmission of audio and video data,
- Package data channel (asynchronous channel), for example to transmit configuration data for a navigation system and to update software in control units.

Requirements

The transmission of multimedia data – both audio and video data – requires a high data rate and also synchronization of the data transfer between source and sink as well as between a number of sinks.

Transfer system

Physical layer

The MOST standard specifies both optical and electrical technologies of the physical layer (transmission layer). The optical transmission layer is widespread and currently uses fiber-optic cables (polymer optical fibers, POF) made of polymethyl methacrylate as its transfer medium. These have a 1 mm core diameter and are used in combination with LEDs and silicon photo diodes as receivers (see Optical fibers/waveguides).

The outstanding feature of MOST 50 is its suitability for the electrical transmission of data. This enables data transfer across unshielded, twisted copper cables (UTP, unshielded twisted pairs). Whereas MOST 25 technology has continued to develop in Europe for a many years and has become established in the Korean market, the Japanese market in particular prefers MOST 50, the second generation of the multimedia standard.

The identification number, for example in the case of MOST 25, stands for a transfer rate of approx. 25 Mbit/s. The exact data rate depends on the sampling rate that the system uses. With a sampling rate of 44.1 kHz, the MOST frame (data frame) is transferred 44,100 times per second; a frame length of 512 bits results in a data rate of 22.58 Mbit/s. For MOST 50, the same sampling rate results in double the data rate, as the frame is 1,024 bits long. Higher data rates of 150 MBit/s (MOST 150) are currently also available.

Special features of MOST 150

In addition to the higher bandwidth of 150 Mbit/s, MOST 150 contains an isochronous transport mechanism to transfer compressed data of HD videos efficiently. MPEG transport streams (MPEG, Moving Picture Experts Group) are transported directly here. With a corresponding MPEG4-based video codec, it is possible to transfer resolutions of up to 1,080 p (1,080 screen lines), as delivered by BluRay players for example. Alongside this, MOST 150 provides an Ethernet channel for the efficient transmission of IP package data (IP = internet protocol).

In contrast to the MAMAC protocol (MOST Asynchronous Medium Access Control) used in the case of MOST 25, the Ethernet channel is able to transfer Ethernet frames. The Ethernet channel transfers unmodified Ethernet data blocks, which means that software stacks and applications from the fields of consumer electronics and IT can be integrated seamlessly in vehicles with much shorter innovation cycles. TCP/IP stacks or protocols that use TCP/IP (TCP = Transmission Control Protocol) can thus communicate via MOST 150 without changes.

The MOST Network Interface Controller (NIC) is a hardware controller that is responsible for control of the physical layer and implements important transfer mechanisms.

Protocol

Data transfer

Data transfer on the MOST bus is organized in data frames that are created by the timing master with a fixed data rate and passed on by the devices in the ring.

Data frames

The timing master usually creates data frames with a clock rate of 44.1 kHz, more rarely also 48 kHz. The size of the data frames thus determines the bus speed of a MOST system. In the case of MOST 25, the size of a data frame is 512 bits (Figure 14). The synchronous and asynchronous areas of MOST 25 jointly use 60 bytes of the data frame. The division between the synchronous channels and the asynchronous channel is determined by the value of the boundary descriptors with a resolution of 4 bytes. The synchronous area must have at least 24 bytes (six stereo channels). This means that between 24 and 60 bytes are permitted for the synchronous area and between zero and 36 bytes are permitted for the asynchronous area. The preamble is used for synchronization; the parity bit is used to recognize bit errors.

Transmission of control messages

The control channel is used to signal device statuses and for the messages required for system administration. So that the control channel does not occupy too much bandwidth per frame, it has been distributed to 16 frames grouped to form a block. Each frame transports two bytes of the channel (Figure 14). In order to ensure recognition of the block start, the preamble of the first frame of a block bears a special bit pattern. In the case of MOST 25, the control channel has a gross bandwidth of 705.6 kbit/s.

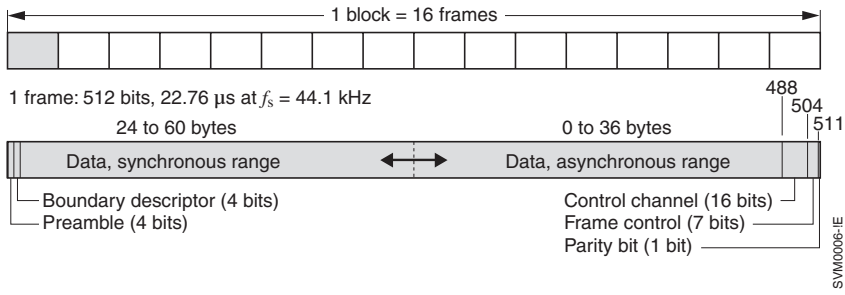
Transmission of multimedia data

The synchronous channels are used for realtime communication of audio and video data, whereby the data interchange is controlled via the corresponding control commands on the control channel. A synchronous channel can be assigned a certain bandwidth, which occurs with a resolution of one byte of a data frame. A stereo audio channel with a resolution of 16 bits, for example, requires four bytes. In the case of MOST 25, depending on the value of the boundary descriptor, a maximum of 60 bytes are available for synchronous channels; this corresponds to 15 stereo audio channels.

Transmission of package data

Data are transferred packet-wise on the asynchronous channel. It is therefore suitable for the transmission of information that has no fixed data rate but requires high data rates at short notice. Examples are the transmission of track information of an MP3 player or a software update.

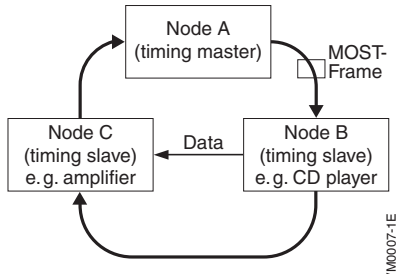
Figure 14: MOST frame structure



In the case of MOST 25, the asynchronous channel has a gross bandwidth of up to 12.7 MBit/s and it currently supports two modes: a slower 48-byte mode in which 48 bytes are available in each package for the net data transmission as well as a 1,014-byte mode that is more complex to implement. In order to ensure reliable transmission and flow control for the typically large asynchronous channel data volumes, an additional transport protocol (Data Link Protocol) is usually deployed; this is implemented in a driver layer at a higher level. This is either the MOST High Protocol (MHP), which has been specially developed for the MOST, or the common TCP/IP protocol that is placed on a corresponding adaptation layer called the MOST Asynchronous Medium Access Control (MAMAC).

Topology
MOST is organized in a ring structure (Figure 15). This is a point-to-multipoint data flow system (i.e. the streaming data have one source and a number of sinks) and all the devices therefore share a joint system frequency that they acquire from the data flow. The devices are thus in phase and can transfer all data synchronously. This makes mechanisms for signal buffering and signal processing superfluous. A certain device acts as "timing master" and generates the data frames used for data transfer and to which the other devices synchronize.

Figure 15: Ring structure of the MOST bus



Addressing
The devices are addressed on the MOST bus via a 16-bit address. Various types of addressing are available: Logical and physical addressing as well as group addressing for simultaneous addressing of a defined group of control units.

Administration functions
The MOST standard defines the management mechanisms (network and connection masters) necessary for the operation of a MOST system. These mechanisms are described below.

Network master
The network master is implemented by a marked device in a MOST system and is responsible for the configuration of the system. In current systems, the network master is usually implemented using the head unit (i.e. control panel) of the infotainment system. This device is often also the timing master at the same time. The other devices of the MOST system are referred to as network slaves in this context.

Connection master
The connection master manages the synchronous connections that exist in a MOST system at any particular point in time.

MOST application layer
For the transfer of control commands, status information, and events, the MOST standard defines a corresponding protocol on the application level. This protocol enables triggering of a certain function of an application interface (i.e. of an FBlock = function block) that is provided by any device within the MOST system.

The protocol for MOST control messages includes the following elements of a control message:
– The address of a device in the MOST system (DeviceID)
– An identifier for an FBlock (FBlockID) implemented by this device and its instance in the MOST system (InstID)
– The identifier for the function called up within the FBlock (FunctionID)
– The type of an operation (OpType).

Function block
A function block (FBlock) defines the interface of a certain application or of a system service. The sinks and sources for multimedia data are each assigned to an FBlock that provides the corresponding functions for their administration. An FBlock can therefore have a number of sources and sinks enumerated via a source and sink number.

An FBlock has functions that deliver information regarding the number and type of sources and sinks that it provides (SyncDataInfo, SourceInfo and SinkInfo). Over and above this, each FBlock with a source has an "Allocate" function that it uses to request a synchronous channel and connect the source to that channel. Accordingly, an FBlock with a sink has a "Connect" function to connect the sink to a certain synchronous channel and a "DisConnect" function to break off this connection.

An FBlock is addressed by means of an 8-bit FBlockID, which specifies the type of the FBlock, and an additional 8-bit InstID.

Function classes
To standardize the way in which functions are defined, the MOST standard specifies a series of function classes for properties. These determine which properties the function has and which operations are permissible.

Applications
As well as defining the lower layers necessary for data transfer, the MOST standard defines the interfaces for typical applications from the area of vehicle infotainment systems, e.g. a CD changer, amplifier or radio tuner.

The FBlocks defined by the MOST Cooperation are summarized in a function catalog.

Standardization
The MOST standard is maintained by the MOST Cooperation, which also publishes the corresponding specifications. The specifications are available on the homepage of the MOST Cooperation [7].

The MOST Corporation was founded in 1998 by BMW, Daimler, Becker Radio, and OASIS Silicon Systems with the aim of standardizing MOST technology.

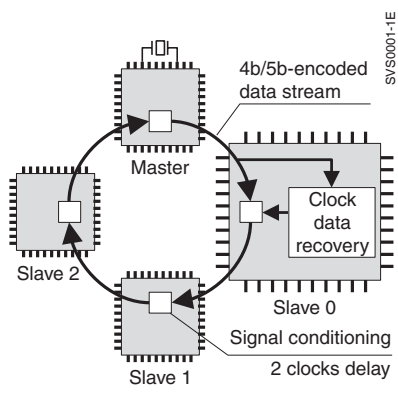
Serial Wire Ring

Internal buses do not leave the ECU. They connect the integrated circuits (ICs) on the board to each other. These buses transmit data from input circuits (e.g. analog-digital converters) to the controller or data to activate output stages from the controller to the output stages. Since these buses are only used inside one ECU, voltage levels, signal type (symmetrical or asymmetrical), impedances, and bit rate can be configured differently, depending on the requirement.

Features
The Serial Wire Ring (SWR) is the modern successor to the Serial Peripheral Interface (SPI). Its advantages are the higher bit rate and only two pins per user as opposed to four pins (CLK, MOSI, MISO, CS) in the SPI. As the price for these advantages the SWR requires a more complex logic in the transmitter and in the receiver, particularly to recover the clock signal from the data stream.

Only two pins per user provide for small housings and less wiring on the board. This is made possible by a ring topology as pictured in Figure 1. This has the additional benefit of the master being able to read back all data and thereby discover transmission errors. This is an important feature for safety-related systems.

Figure 1: Topology of the Serial Wire Ring



Topology
The ring consists exclusively of point-to-point connections, providing for high data rates. Each connection transmits the data from precisely one source to precisely one destination in only one direction (Figure 1):

Master → Slave 0 → Slave 1 → … → Slave $N-1$ → Master.

The ring contains one master and 1 to 16 slaves. Since all are similarly connected to each other in the ring, it is optionally possible for alternately a different module in the ring to assume the role of master.

Each user in the ring forwards the received signal with two clocks delay to the next user. The level of the signal is thereby conditioned. Even small time fluctuations of the signal edges (jitter) are reduced with the aid of clock data recovery (CDR). In this way, different signal levels can be realized within a ring.

4b/5b encoding of data
Clock and data are jointly transmitted via a line in the ring. The receiver extracts the data and the clock from this signal by means of clock data recovery (CDR). To enable the clock to be recovered even in the event of many consecutive “0” or “1”, a special 4b/5b encoding as depicted in Table 1 is used: Every fourth bit is inverted and then sent again. With this encoding there is a change from “0” to “1” or vice versa after five clocks at the latest. The drawback is that five bits have to be sent over the bus to transmit four data bits.

Table 1: 4b/5b code
The lowest-value bit is transmitted first.

0000 → 00001	0001 → 00010
1000 → 10001	1001 → 10010
0100 → 01001	0101 → 01010
1100 → 11001	1101 → 11010
0010 → 00101	0011 → 00110
1010 → 10101	1011 → 10110
0110 → 01101	0111 → 01110
1110 → 11101	1111 → 11110

Transmission is performed according to the following scheme: The lowest-value bit is always transmitted first and the highest-value bit is transmitted last.

Clock rate
The clock rate is not tied down to specific values, but can be chosen to suit the required data rate and latency. The lower limit is 1 MHz, the upper limit is determined by the slowest user and the line routing.

Signal transmission from precisely one source to precisely one destination facilitates higher data rates than with a branched line such that the Serial Wire Ring can be operated at 10 to 100 times the clock frequency of the Serial Peripheral Interface. The demand for higher speed mentioned at the start is thus fulfilled.

Data transmission
The interframe symbol and the data frame or the interrupt frame are transmitted alternately by the master on the data line (2).

Addressing
The slaves are addressed via an address field in the data frame. Each slave receives its address automatically via its position in the ring during the initialization of the ring. The first slave receives the address 0, the last of N slaves receives the address $N-1$.

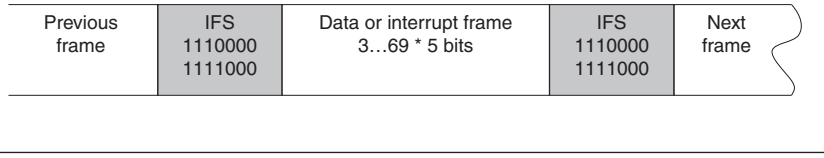
Data frame
The data frame serves to transmit user data between master and slave. With the same data frame the master can also send data to several slaves and receive data from several slaves. At the same time it is possible to transmit data from slave to slave, but only in the circulation direction of the ring.

The master continuously sends frames through the ring. If no data have to be transmitted, it sends “zero frames”.

Interrupt frame
Interrupt frames are regularly sent by the master within a fixed time interval. Each slave which has reserved one or more bits in the interrupt frame can use such a bit to trigger a specific reaction in the master. This reaction usually consists of delivering or collecting data.

Interframe symbol
To enable each user in the ring to detect the start of a frame, the interframe symbol (IFS, Figure 2) is situated between two frames. It is 14 bits long and differs from each possible 14-bit sequence of the 4b/5b-encoded signals in at least two bits. This ensures that a 1-bit error does not lead to the appearance of an interframe symbol in the regular 4b/5b data stream. Each user in the ring compares the last 14 received bits with the interframe symbol. If they match, the user begins to evaluate the frames that now follow. In the event of errors, it aborts the evaluation and waits for the next interframe symbol.

Figure 2: Structure of a data transmission
IFS Interframe symbol.



Data frame

Identifier

Figure 3 shows a data frame in detail. The leading 1 identifies the frame as a data frame. A 01 would indicate a zero frame and a 0010 an interrupt frame.

Data frame counter

The data frame counter (DFC) is a 2-bit counter which the master increments before it sends a new data frame. This is important in the event of a transmission error that the master notices when reading back. It then repeats the data frame with the same DFC. If the error has occurred on the way from the master to the slave, the slave now receives the data frame for the first time and evaluates it. If the error has occurred on the way from the slave to the master, the slave now receives the data frame for the second time, detects this by the unchanged DFC, and now does not evaluate it for a second time.

Slave address

The slave address SLADR addresses the slave. The primary addresses 0 to $N-1$ address precisely one slave at the corresponding position in the ring. Higher values up to 31 are used for group addresses. While the ring is being initialized, the master can instruct one or more slaves to execute a particular operation automatically if a certain group address appears, for example to output bits 12 to 16 to five output stages. In this way, the master can carry out complex data transfers later without further instructions and thus very quickly simply by using the associated group address. The participating slaves then "know" what is to be done.

Parity bit

PAR is a parity bit for the first eight bits. It is sufficient in order to detect 100 % of all single-bit errors. Double-bit errors are detected by the master when reading back and lead to a bus reset. Therefore they do not have to be detected by PAR. This procedure is possible because double- and multiple-bit errors are very rare.

Data field

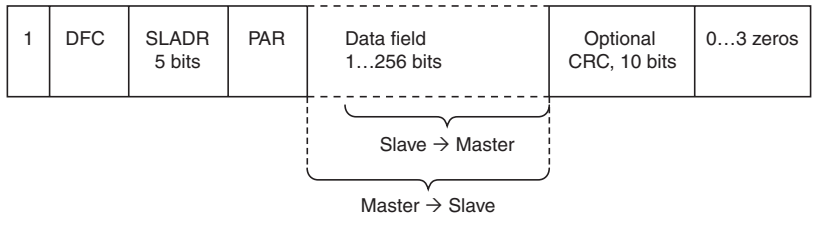
The actual data field comes after PAR. It can contain any number of bits from 1 to 256.

Primary address

In the event of a primary address, i.e. if the master only addresses this slave, the slave processes the entire data field and if necessary places an answer at the end of the data field. In the process it overwrites part of the data field so that the master can no longer read back its message to the slave. If this is not permitted for security reasons or if the slave's answer is already longer than the master's message, the master must configure the data field bigger and for example append zeros. The slave then overwrites only this areas and no old data are lost. However, this gives rise to a larger data field, i.e. a longer transmission time.

Figure 3: Data frame

DFC Data frame counter,
SLADR Slave address,
PAR Parity bit,
CRC Cyclic redundancy check.



Group address
In the event of a group address, each slave can, according to its previously learned instructions, evaluate and change any bits of the data field.

Cyclic redundancy check
The concluding 10-bit CRC (cyclic redundancy check) is supported by slaves for which this is useful. There are also simple slaves which do not support the CRC in order to save costs. The CRC is required to back up critical data on their way from the slave to the master. Because, while the master can detect all the other transmission errors when reading back, this is not possible with the answer data from a slave. The slaves which belong to a group address must all jointly use the CRC, or not.

Filler bits
Because of the 4b/5b encoding the data frame must always contain a multiple of four bits. It may therefore be necessary finally to append up to three more zeros in order to reach the next multiple higher of four.

Sequence control in the master
As transmission rates increase, it is becoming increasing more difficult for a CPU to make available the transmission data in good time via a simple send and receive register and to collect the received data fast enough. Therefore the SWR master automatically accesses the memory. This memory also contains one or more chained lists with instructions as to which transactions the master is to execute (queues).

Once activated, the master works through a queue instruction by instruction down to the last instruction or to an instruction with a set stop bit. A software command, an external trigger (e.g. by a timer) or the already described prompt by a slave by means of interrupt frame can reactivate the queue.

The end of a queue can refer back to the start and thereby form a closed loop. In this way, the same instructions can be executed again and again. If an instruction is disrupted by a transmission error, the master repeats it automatically.

Each instruction contains a priority. This does not play a role as long as only one queue is active. However, as soon as several queues are active and therefore several instructions are due to be executed at the same time, the master selects the instruction with the highest priority.

References
[1] ISO 11898: Road vehicles – Controller area network (CAN).
Part 1: Data link layer and physical signalling.
Part 2: High-speed medium access unit.
Part 3: Low-speed, fault-tolerant, medium-dependent interface.
Part 4: Time-triggered communication.
Part 5: High-speed medium access unit with low-power mode.
Part 6: High-speed medium access unit with selective wake-up functionality.
[2] SAE J 1939: Serial Control and Communications Heavy Duty Vehicle Network.
[3] ISO 17458: Road vehicles – FlexRay communications system.
Part 1: General information and use case definition.
Part 2: Data link layer specification.
Part 3: Data link layer conformance test specification.
Part 4: Electrical physical layer specification.
Part 5: Electrical physical layer conformance test specification.
[4] http://www.lin-subbus.org/.
[5] http://www. IEEE802.org/.
[6] http://www.psi5.org/.
[7] http://www.mostcooperation.com.

Automotive sensors

Automotive applications

Sensors and actuators form as the periphery the interface between the vehicle with its complex drive, brake, chassis-system, and body functions and the digital electronic control unit (ECU) as the processing unit. The sensors detect physical quantities and convert them into electrical quantities that can be processed by the ECU.

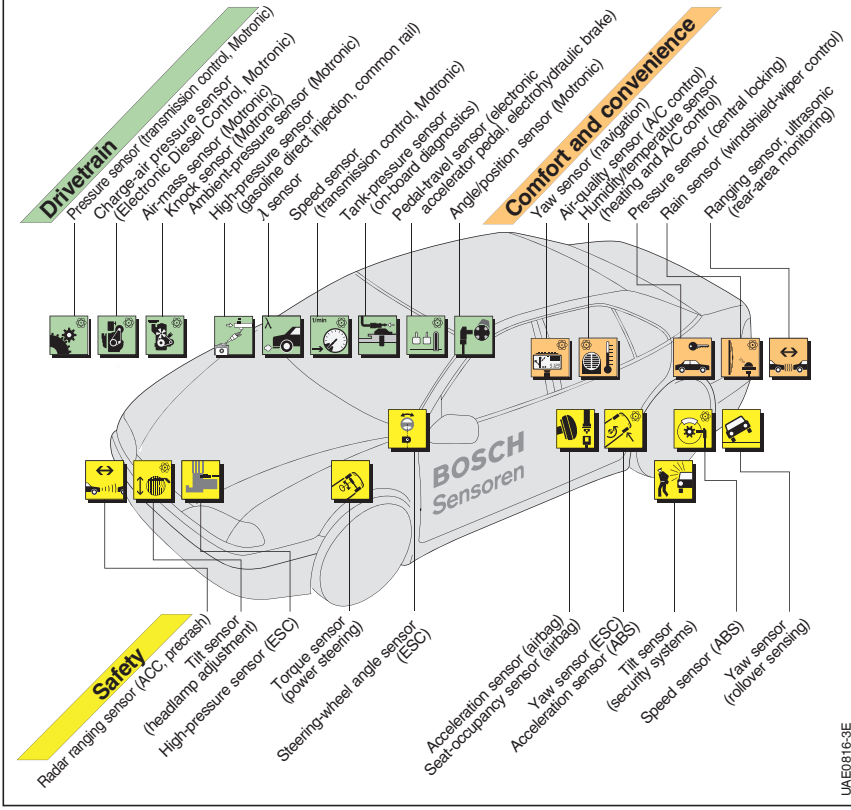
Classification

Classic sensors

Sensors were introduced in motor vehicles with the emergence of electronic control for fuel injection and Ignition. Speed sensors to acquire the engine speed and pressure sensors to measure the intake-manifold pressure were required for this purpose. With the antilock braking system (ABS) it was necessary to acquire the wheel speed. Wheel-speed sensors which operate according to the same measuring principle as the engine-speed sensors were used for this purpose. The further development of these speed sensors led from the inductive sensor to the active sensor with an improvement in measured-value acquisition.

The constantly rising complexity of the electronic systems in motor vehicles led to

Figure 1: Automotive sensors (examples)
The applications of the respective sensors are specified in parentheses.



an increase in sensors. Figure 1 provides an overview of where and for which purposes such sensors are used.

Sensors for driver-assistance systems
A completely new type of sensor was introduced with the ultrasonic sensors for the parking aid. These sensors as well as radar and video sensors sense the vehicle's environment or surroundings and thereby facilitate a wealth of driver-assistance functions (see Driver-assistance systems). These sensors are complex systems and cannot be compared with the classic sensors.

Main requirements
Operating conditions
Automotive sensors are sometimes exposed to extreme conditions in their installation locations. These are mechanical stresses (e.g., vibrations and impacts), climatic influences (e.g. low and extremely high temperatures, moisture), chemical influences (e.g., splash water, salt spray, fuel, engine oil, battery acid), and electromagnetic influences (e.g., high-frequency interference, mains-borne interference pulses, overvoltage). The degree of stress to which a sensor is subjected is determined by the operating conditions at its installation location.

Reliability
According to application and technical requirements, automotive sensors are assigned to one of three reliability classes:
– Steering, brakes, passenger protection,
– Engine, drivetrain, suspension, tires,
– Comfort and convenience, information, diagnosis, theft deterrence.

Size
The constantly increasing number of electronic systems in vehicles on the one hand and the ever more compact shape of vehicles on the other hand force engineers to adopt extremely small designs. In addition, the pressure of increased fuel economy necessitates a consistent reduction of vehicle weight. Various miniaturization concepts are employed to achieve compact sensor dimensions:
– Substrate and hybrid technologies (e.g., for strain-dependent resistors, thermistors, and magnetoresistors),
– Semiconductor technology (e.g., for Hall-effect crankshaft sensors),
– Surface and bulk micromechanics (e.g., for silicon pressure, acceleration, and yaw sensors),
– Microsystem technology (integration of micromechanical or microoptical components with microelectronic circuits in a complex system).

Manufacturing costs
Electronic systems in vehicles contain up to 150 sensors. This abundance of sensors forces engineers to radically reduce costs compared with other areas of use. Extensively automated manufacturing processes working at great efficiency levels are used. In other words, each process step is always performed simultaneously for a larger number of sensors. The manufacture of semiconductor sensors typically involves the simultaneous integration of 100 to 1,000 sensors on one silicon wafer. The automotive industry's huge demand for sensors has set new standards here.

Accuracy requirements
The accuracy requirements for automotive sensors are lower than for sensors for the process industry for example. The permissible deviations are generally greater than 1 % of the upper range value of the measurement range. The inevitable influences of aging must also be taken into consideration here.

However, ever more sophisticated and complex systems call for higher levels of accuracy. This can be achieved to a certain extent by reducing the manufacturing tolerances and by improving the adjustment and compensation technologies. "Integrated sensors" provide for a significant advance here.

Basic principles of classic sensors

Function

Sensors convert a physical or a chemical (usually non-electrical) quantity Φ into an electrical quantity E. This is often also done via further, non-electrical intermediate stages. The sensor data are processed in a digital electronic control unit (ECU) (Figure 2). An adapter circuit is generally used to convert the sensor's signals into the standardized form required by the ECU. In addition, system operation can be influenced by sensor information from other processing elements for example via bus systems and/or by driver-operated switches.

Sensor classification

Sensors can be classified according to very different standpoints. With regard to their use in motor vehicles, they can be categorized as follows.

Assignment and application

– Functional sensors (e.g., pressure sensor, temperature sensor, air-mass sensor), primarily for open- and closed-loop control functions.
– Sensors for safety (passenger protection: airbag and driving-dynamics control) and protection (theft deterrence).
– Sensors for vehicle monitoring (onboard diagnostics, consumption and wear quantities) and for providing information to driver and passengers.

Type of characteristic curve

– Continuous linear characteristic curves (Figure 3a) are used particularly for control functions across a broad measurement range. Linear characteristic curves furthermore have the advantage that they are easy to check and adjust.

Figure 3: Types of characteristic curve
S Output signal, X Measured variable.
a) Continuous, linear,
b) Continuous, nonlinear,
c) Discontinuous, multistage,
d) Discontinuous, dual-stage with hysteresis.

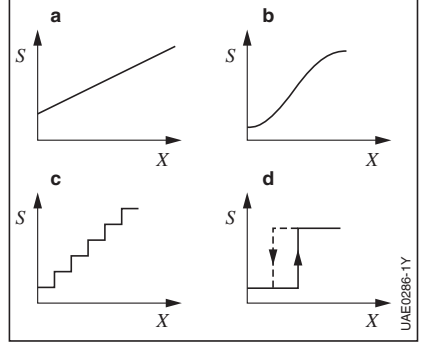


Figure 2: Automotive sensor
Φ Physical quantity,
E Electrical quantity,
Z Disturbance.
AC Actuator, AD Adapter circuit, DPL Display,
MS Measuring sensor, SA Switch,
MS Measuring sensor, SA Switch,
SE Sensor, ECU Electronic control unit,
FCE Final controlling element, D Driver.

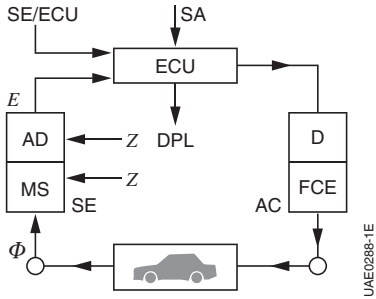
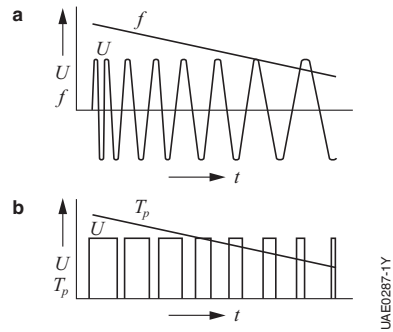


Figure 4: Signal shapes (examples)
a) Frequency f,
b) Pulse duration T_p,
U Output signal, t Time.



– Continuous nonlinear characteristic curves (Figures 3b) are often used for closed-loop control of a measured variable within a very narrow measurement range (e.g., control of the air/fuel ratio at $\lambda = 1$, control of the compression level).
– Discontinuous dual-stage characteristic curves, with hysteresis in some cases (Figure 3d) are used to monitor limit values where an easy remedy is possible when they are reached. If remedy is more difficult, early warning can be effected at an earlier stage by means of multiple stages (Figure 3c).

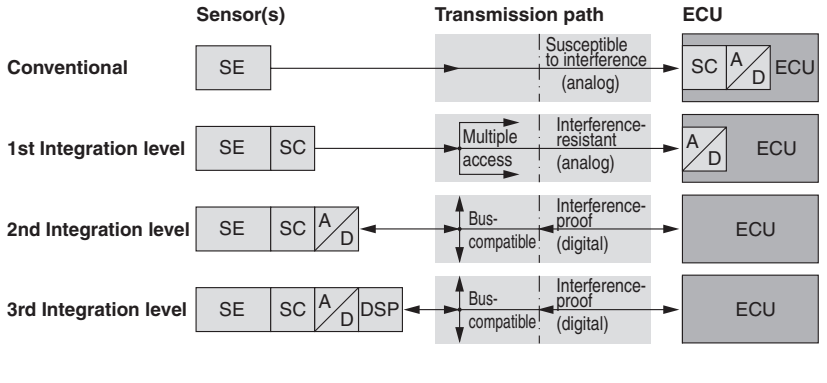
Type of output signal
Analog output signal (Figure 4):
– Current or voltage, or corresponding amplitude,
– Frequency or period duration,
– Pulse duration or pulse duty factor.

Discrete output signal:
– dual stage (binary encoded),
– multistage – irregular graduation (analog encoded),
– multistage – equidistant (analog or digital encoded).

Continuous signals are constantly available at the sensor output, discontinuous signals are available only at discrete times. A bit-serially output signal is inevitably discontinuous.

Integrated sensors
Systems range from hybrid and monolithic integrated sensors and electronic signal-processing circuits at the measuring point, through to complex digital circuits, such as analog-digital converters and microcomputers, for complete utilization of the sensor's inherent accuracy "intelligent" sensors, Figure 5). These systems offer the following benefits and options:
– Reduction of load on the ECU,
– Uniform, flexible, bus-compatible interface,
– Multiple application of sensors,
– Multisensor designs,
– Use of small measuring effects and high-frequency measuring effects (by local amplification and demodulation), and
– Correction of sensor deviations at the measuring point, and common calibration and compensation of sensor and circuit, are simplified and improved by storage of the individual correction information in semiconductor memory chips (e.g., PROM).

Figure 5: Sensor integration levels
SE Sensor, SC Signal conditioning (analog), AD Analog-digital converter, ECU Electronic control unit, DSP Digital signal processing.



Micromechanics

Applications

Micromechanics is defined as the application of semiconductor technology in the production of mechanical components from semiconductor materials (usually silicon). This type of application exploits both the semiconducting and the mechanical properties of silicon. The first micro-mechanical silicon pressure sensors were installed in motor vehicles at the beginning of the 1980s. Typical mechanical dimensions can extend into the micrometer range.

The mechanical properties of silicon (e.g. strength, hardness and modulus of elasticity, see Table 1) can be compared to those of steel. However, silicon is significantly lighter and has greater thermal conductivity than steel. Single-crystal silicon wafers are used with almost perfect mechanical properties. Hysteresis and creepage are negligible. Due to the brittleness of the single-crystal material, the stress-strain curve has no plastic range; the material ruptures when the elastic range is exceeded.

Two methods of manufacturing micromechanical structures in silicon have established themselves: volume micromechanics (VMM) and surface micromechanics (SMM). Both methods use the standard procedures of microelectronics (e.g. epitaxial growth, oxidation, diffusion, and photolithography) together with some additional special procedures [1].

Volume micromechanics

The silicon wafer material is processed at the required depth using anisotropic (alkaline) etching with or without an electrochemical etching stop. From the rear, the material is removed from inside the silicon layer, where there is no etching mask resting on top (Figure 5). Using this method, very small membranes can be produced with typical thicknesses of between 5 and 50 µm, as well as openings, beams and webs as are needed for instance in pressure and acceleration sensors.

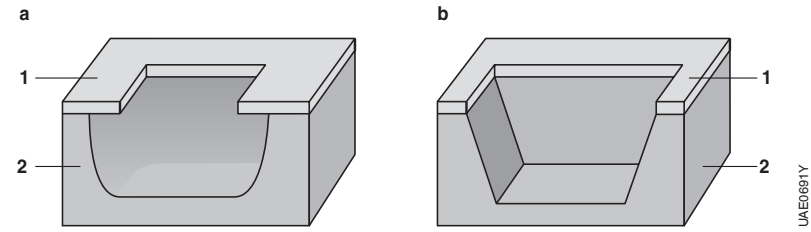
The problem when etching with alkaline media lies in the fact that the walls run inwards at an angle. In order to be able to effect high-precision etching at depth with vertical walls, it was necessary to develop a new process. Here Bosch made a breakthrough with the development of the DRIE process (Deep Reactive Ion Etching), which has in the meantime come to be known generally in the industry as the “Bosch process”. This involves the creation in a special gas-phase reactor of alternating gases and accompanying conditions for an etching

Table 1: Mechanical properties of silicon

Quantity	Unit	Silicon	Steel (max.)	Stainless steel
Tensile load	10^5 N/cm^2	7.0	4.2	2.1
Knoop hardness	kg/mm^2	850	1,500	660
Modulus of elasticity	10^7 N/cm^2	1.9	2.1	2.0
Density	g/cm^3	2.3	7.9	7.9
Thermal conductivity	W/cm · K	1.57	0.97	0.33
Thermal expansion	10^{-6}/K	2.3	12.0	17.3

Figure 5: Electrochemical etching

a) Isotropic (in acid etching media), b) Anisotropic (in alkaline etching media).
1 Etching mask (e.g. oxide or nitride), 2 Silicon.



stage and then a passivation stage. This etching and the subsequent passivation of the walls are set such that very precise vertical walls are created. This process can be used for both volume and surface micromechanics.

Surface micromechanics
In contrast to volume micromechanics, surface micromechanics merely uses the silicon wafer as the substrate. Moving structures are usually formed from polycrystalline silicon layers which, similarly to a manufacturing process for integrated circuits, are deposited on the surface of the silicon by epitaxial growth.

When a surface-micromechanical component is made, a sacrificial layer of silicon oxide is first applied on the wafer and structured with standard semiconductor processes, i.e. partly removed again specifically (Figure 6a). An approx. 10 µm thick polysilicon layer (epipoly layer) is then applied at high temperatures in an epitaxial reactor (Figure 6b) and whose desired structure is anisotropically, i.e. vertically, etched with the aid of a lacquer mask (deep etching or trenching, Figure 6c). The vertical side walls are obtained with the Bosch process by alternating etching and passivation cycles. After an etching cycle, the etched sidewall section is provided during passivation with a polymer as protection so that it is not attacked during the subsequent etching. Vertical side walls with a high representation accuracy are created in this way. In the last process stage (Figure 6d) the sacrificial layer beneath the polysilicon layer is removed with gaseous hydrogen fluoride in order to expose the structures (Figure 7).

Among other things, surface micromechanics is used in the manufacture of capacitive acceleration sensors for airbag systems and in the manufacture of yaw sensors for applications in driving-dynamics control and in rollover sensing.

APSM process
The "Advanced Porous Silicon Membrane" process (APSM process) uses a completely different surface-micromechanical process sequence. It uses the properties of porous silicon to create

Figure 6: Process stages in surface micromechanics
a) Depositing and structuring the sacrificial layer,
b) Depositing polysilicon,
c) Structuring polysilicon by deep etching,
d) Removing the sacrificial layer and so producing freely moving structures on the surface.
1 Silicon,
2 Oxide layer (sacrificial layer),
3 Polysilicon layer ("epipoly").

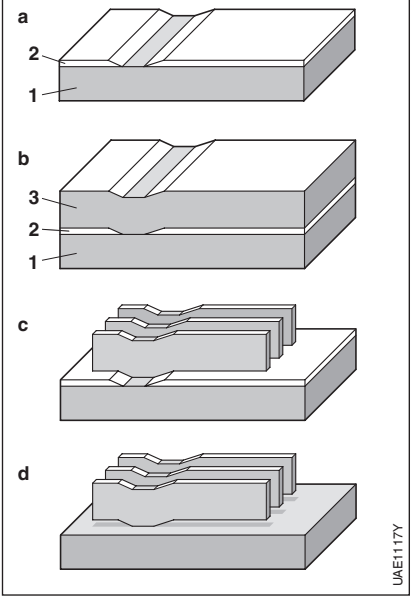
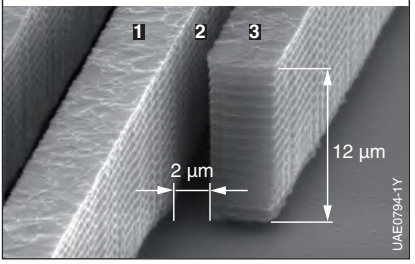


Figure 7: Structure of a surface-micromechanical sensor
Recording with scanning electron microscope.
1 Fixed electrode,
2 Gap,
3 Spring electrode.



beneath a monocrystalline membrane an exactly defined cavity in which a vacuum is enclosed.

The core of the APSM process is porous silicon. This can be manufactured selectively and on a locally limited basis in p-doped silicon in an electrochemical anodizing process in hydrofluoric acid. In this process some of the silicon is dissolved out of the crystal to leave a porous, sponge-like silicon shell, the "porous silicon".

Porous silicon in turn can be rearranged at high temperatures. The silicon shell dissolves and forms under suitable conditions a thin membrane on the surface. A cavity forms below this membrane (Figure 8). This thin membrane can be enlarged to the target thickness by means of epitaxial growth.

The monocrystalline epitaxial layer is used for example as the diaphragm of a pressure sensor. Evaluating-circuit elements in the epitaxial layer outside the diaphragm provide for a high-precision, small and inexpensive pressure sensor.

Modern pressure sensors, e.g. barometric pressure sensors for engine-management systems, are manufactured using the APSM process.

Wafer bonding

In addition to structuring the silicon, joining two wafers represents another essential task in micromechanical production engineering. Joining technology is required, for example, in order to hermetically seal a reference-vacuum chamber (e.g. for pressure sensors), to protect sensitive structures by applying caps (e.g. in acceleration and yaw sensors, Figure 9), or to join the silicon wafer with intermediate layers which minimize the thermal and mechanical stresses (e.g. glass base on pressure sensors).

In the case of anodic bonding, a Pyrex-glass wafer is joined to a silicon wafer at a voltage of some 100 V and a temperature of approx. 400 °C (Figure 10). A strong electrostatic attraction and an electrochemical reaction (anodic oxidation) result in a permanent hermetic bond between the glass and the silicon.

In the case of seal-glass bonding, two silicon wafers are contacted by way of a glass-solder layer applied in a screen-printing process at approx. 400 °C and under the exertion of pressure. The glass solder melts at this temperature and produces a hermetically sealed bond with the silicon.

References for Sensors – Basic principles

[1] U. Hilleringmann: Mikrosystemtechnik – Prozessschritte, Technologien, Anwendungen. B.G. Teubner-Verlag, Wiesbaden 2006.

Figure 8: Creating exact vacuum caverns in silicon with the APSM process

1 Silicon,
2 Cavity, created from porous silicon by the APSM process,
3 Membrane,
4 Evaluating circuit.

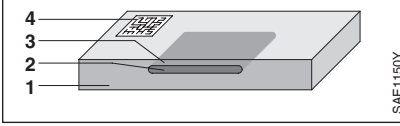


Figure 9: Thin-layer cap for hermetically sealing e.g. an acceleration sensor

1 Silicon,
2 Freely moving structure,
3 Cap.

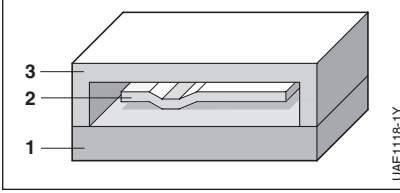
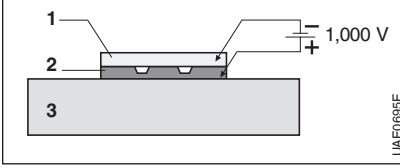


Figure 10: Anodic wafer bonding

1 Pyrex glass,
2 Silicon,
3 Heating plate ($T \approx 400$ °C).



Position and angular-position sensors

Measured variables

These sensors record one- or multidimensional displacement and angle positions (translational and rotatory variables) of the most varied types and ranges. Examples of such measured variables are:

- Accelerator-pedal position for recording the torque request for engine management (driver command),
- Throttle-valve position for controlling the throttle valve,
- Fuel-tank level,
- Steering-wheel angle,
- Position of the transmission selector lever for electronic transmission control
- Seat position,
- Mirror position,
- Control-rack position in the diesel in-line fuel-injection pump,
- Travel of clutch servo unit,
- Brake-pedal position,
- Tilt angle.

In this sector of applications, activities have long since been directed at changing over to proximity or non-contacting sensor principles. Such sensors are wear-free and thus have a longer service life as well as being more reliable. The costs involved though, often force vehicle manufacturers to retain the "wiper-type" sensor principle, and such sensors still perform efficiently enough for many measurement purposes.

In practice, incremental sensor systems are also often referred to as angular-position (or angle-of-rotation) sensors, even when they are used for measuring rotational speed. These sensors are in reality not angular-position (or angle-of-rotation) sensors. Since the increments (steps with which a given quantity increases) which have to be measured with these sensors in order to measure the deflection angle must first be counted with the correct preceding sign (in other words, added), there is the possibility of permanent interference.

Wiper potentiometer

Measurement principle

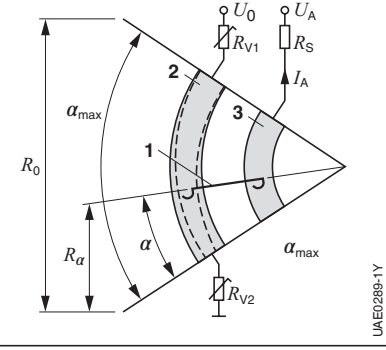
The wiper potentiometer – usually designed as an angular-position sensor – measures travel by exploiting the proportional relationship between the length of a wire or film resistor (conductor track made of cermet or "conductive plastic") and its electrical resistance (Figure 1). At present, this is the lowest-priced travel and angular-position sensor.

To protect against overloading, supply voltage is usually applied to the measurement track through series resistors R_V. These resistors can also be used for calibration of zero point and curve slope. The shape of the contour across the width of the measurement track (including that of individual sections) influences the shape of the characteristic curve.

The standard wiper connection is furnished by a contact track consisting of the same material mounted on a low-resistance substrate. Wear and measurement distortions can be avoided by reducing the current at the pick-off (I_A < 1 mA) and sealing the unit against dust and liquid. Also essential to minimizing wear is to have an optimum friction pairing of wiper

Figure 1: Wiper potentiometer

1 Wiper,
2 Potentiometer track (measurement track),
3 Contact track.
U_0 Supply voltage,
U_A Measurement voltage,
I_A Wiper current,
R_0 Resistance of measurement track,
R_α Resistance of part of measurement track,
R_V Series resistors, R_S Protective resistor,
α Measurement angle.



and potentiometer track; here wipers can have a spoon or scraper shape and have a single or multiple design, even in the shape of a broom.

The wiper potentiometer together with further resistors form a voltage divider (Figure 2). The picked-off voltage is a measure of the wiper position.

Advantages of the potentiometer are, for example, the low costs, the simple and clear design, the large measuring effect (measuring stroke corresponds to the supply voltage), the wide operating-temperature range up to 250 °C, and the high degree of accuracy (better than 1 % of the upper range value of the measurement range).

Disadvantages of the potentiometer are, for example, mechanical wear caused by abrasion, measuring errors caused by abraded remnants, wiper lift-off in the event of vibrations, and the limited miniaturization capability.

Wiper-potentiometer applications

Throttle-valve angle sensor

Mounted on the throttle valve in a gasoline engines is a throttle-valve sensor (Figure 3), which signals the position of the throttle valve to the engine-management system. The throttle-valve angle sensor has a redundant design for diagnostic purposes (Figure 2). This makes available two independent signals which can be checked for plausibility against each other. Redundant measurement enables faults to be detected.

Accelerator-pedal sensor

The accelerator-pedal (pedal-travel) sensor senses the travel or the angle of the pressed accelerator pedal. The engine-management system interprets this value as a torque request from the driver. The accelerator-pedal sensor is integrated together with the accelerator pedal in the accelerator-pedal module. With these units that are ready for installation there is no need for adjustment on the vehicle.

With the potentiometric sensor design the measurement signal is generated by a potentiometer. The engine ECU converts the measured voltage using the stored sensor curve into the covered pedal travel or into the accelerator pedal's angular position.

Like the throttle-valve sensor the accelerator-pedal sensor has a redundant design. One sensor version operates with a second potentiometer that always delivers half the voltage of the first potentiometer at all operating points. Two independent

Figure 2: Electrical circuit of throttle-valve sensor

1 Throttle valve,
2 Throttle-valve sensor.
U_A Measurement voltages,
U_V Operating voltage,
R_1, R_2 Potentiometer tracks 1 and 2,
R_3, R_4 Trimming resistors,
R_5, R_6 Protective resistors.

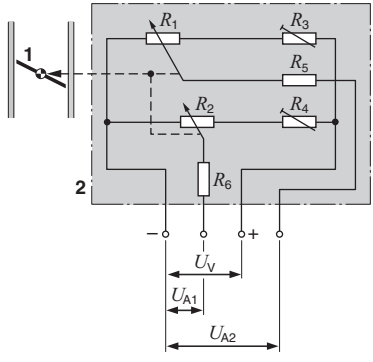
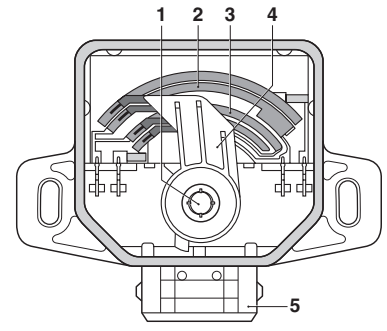


Figure 3: Throttle-valve sensor

1 Throttle-valve shaft,
2 Potentiometer track 1,
3 Potentiometer track 2,
4 Wiper arm with wipers,
5 Electrical connection.



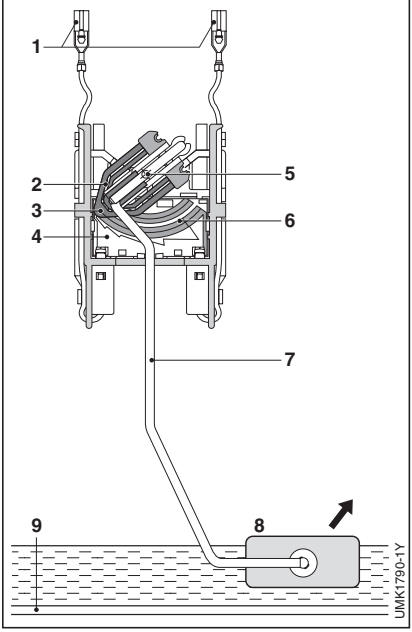
signals are thus available for the purpose of fault detection. Sensor versions with low-idle switches were also customary in the past; the state of these sensors had to be plausible with the measurement signal of the potentiometer.

A further switch for detecting a kick-down can be integrated in the accelerator-pedal module of vehicles with automatic transmissions. Alternatively, this signal can be derived from the rate of change of the potentiometer voltage or generated when a threshold value is exceeded.

Fuel-level sensor

The level in the fuel tank is transmitted by the float via the float lever to the potentiometer (Figure 4). Firstly, the prepared measured value is transmitted for display in the instrument cluster; secondly, the fuel level is used to measure fuel consumption (e.g. for displaying the range calculated by the on-board computer).

Figure 4: Potentiometric fuel-level sensor
1 Electrical connections,
2 Wiper spring,
3 Contact rivet,
4 Resistor board,
5 Bearing pin,
6 Twin contact,
7 Float lever,
8 Float,
9 Base of fuel tank.



Magnetically inductive sensors

Of all the sensors which perform non-contacting position measurement, magnetic sensors are particularly non-susceptible to interference and robust. This applies in particular to those principles relying on alternating current, in other words magnetically inductive principles. Compared with a micromechanical sensor though, the coil configuration needed here requires far more space. This means therefore that there is no favorable possibility of redundant (parallel measurement) design.

Measurement principle, short-circuiting-ring sensors

Of all the established principles of magnetically inductive sensors it is mainly short-circuiting-ring sensors that have up to now been used by Bosch in motor vehicles. These sensors sense for example the control-rack position in diesel in-line fuel-injection pumps [1]. However, these are no longer used in a macromechanical design form for new developments.

Measurement principle, eddy-current sensors

When an electrically conducting disk (e.g. made of aluminum or copper) approaches a (usually air-core) coil supplied with high-frequency alternating current, eddy currents are created in the disk by the increasing magnetic coupling as a function of the distance (measurement travel s). This influences both the equivalent resistance and the inductance of the coil. The disk thus acts as a damper disk.

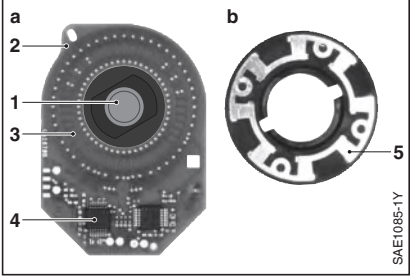
Both the damping effect (on account of the equivalent resistance) and the field-displacement effect (on account of the inductance) can be used to convert the measuring effect into an electrical output voltage. The suitable choice is in the first case for example an oscillator of variable amplitude and in the second case an oscillator of variable frequency or a constantly powered, inductive voltage divider.

A high operating frequency is required due to the low coil inductance. This necessitates a direct assignment of electronic circuitry to the sensor.

Application: Position sensor
for transmission control
In automatic transmissions position sensors sense the position of final controlling elements (e.g. selector-lever shaft, selector valve or parking-lock cylinder). Because of the complex requirements from the different transmission topologies and the space and functional requirements different physical measurement principles (Hall, AMR, GMR and eddy-current principles) and types (linear and rotatory sensing) are used.

In the eddy-current sensor the rotor rotates with the selector-lever shaft (Figure 5). A return coil is mounted on the rotor. The fixed sensor board contains redundant transmitter and receiver coils with the associated evaluation electronics. The transmitter coil induces in the return coil an eddy current opposing its cause whose magnetic field induces a voltage in the receiver coil. The geometries of the return coil and the receiver coils are matched to each other in such a way that continuously variable rotor positions can be sensed. The rotatory position sensor thus determines the selector-lever positions P, R, N, D, 4, 3, and 2.

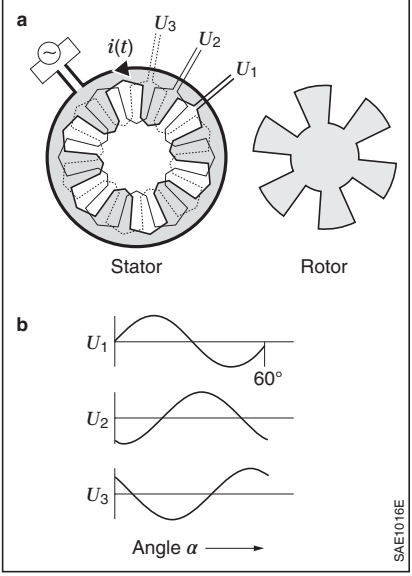
Figure 5: Position sensor using eddy-current principle
a) Sensor board with protective circuit,
b) Rotor.
1 Selector-lever shaft,
2 Sensor board,
3 Redundant transmitter and receiver coils,
4 Redundant evaluation electronics,
5 Rotor with return coil.



For safety reasons the sensor generates two independent, opposite signals which are checked for plausibility against each other.

Sensors with rotating alternating fields
Two- or multipole alternating-field structures can be arranged either in a circle or linearly with coils energized with alternating current of angular frequency Ω or coil-like configurations (such as meander conductor-track structures) (Figure 6). Compared with a usually fixed set of receiver coils, which have the same pole pitch, these pole structures with fixed pole pitch can be displaced by the movement of the system to be measured – whether it be rotatory or translational in nature. Here, the amplitudes of the receiver signals (U_1, U_2, U_3, etc.) change sinusoidally with the movement. If the receiver coils are offset against each other by a certain part of the pole pitch T (e.g. T/4 or T/3), the sine shape is phase-displaced in each case by a corresponding angle (e.g. by 90° or 120°). After rectification, the angle

Figure 6: Sensor with rotating alternating fields
a) Schematic layout,
b) Output signals.



of rotation to be measured can be calculated with great accuracy from these voltages. This is the mode of operation of the sensors referred to in classical metrology as synchro, resolver, or inductosyn processes, and preferably configured as angular-position encoders.

The angular-position sensor shown in Figure 6 most resembles the inductosyn process. This is a sensor with a six-figure pole structure (n = 6), which in electrical terms converts an angle of rotation of φ = 60° into a phase displacement of signal amplitudes of α = 360 °. All the necessary conductor-track structures are at least in the case of the fixed part (stator) situated on multi-layer board material. The rotor part can also if necessary be designed as a stamping, either self-supported or mounted on plastic carriers (by hot stamping).

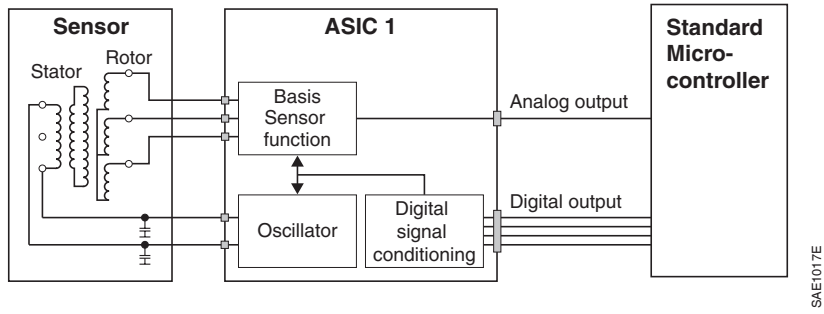
Situated on the stator is a circular conductor-track loop which, independently of the angle of rotation, induces at an operating frequency of 20 MHz an eddy current in a self-contained meander loop of roughly equal outside diameter on the rotor. Like the excitation loop, this eddy current creates a secondary magnetic field which overlaps the excitation field such that it attempts to damp it. If there were only a circular conductor track congruent to the stator loop on the rotor instead of the meander, this would simply extinguish the primary field to the greatest possible extent. However, the meander structure creates a resulting multipole field, which can be rotated with the rotor and whose total flux is also virtually zero.

This multipole alternating field is sensed by concentric receiver coils (meanders) of identical shape likewise located on the stator. These coils or meanders are displaced within a pole pitch (of e.g. 60°) by 1/3 in each case, i.e. electrically in their signal amplitude by 120° in each case (Figure 6b). However, the receiver coils stretch over all the n pole pairs (series connection) and utilize the sum total of all the pole fields.

Figure 7 shows the receiver coils star-connected. For the purpose of determining the electrical phase angle α (or the mechanical angle of rotation), the coil signals are routed to an ASIC, which performs the necessary rectification, selection, and ratio formation. The ASIC receives the necessary digital control signals from a nearby microcontroller. A different version of an ASIC is, however, also able to operate the sensor completely independently (stand-alone). The ASICs also permit an end-of-line adjustment of the mechanical and electrical tolerances in production.

For applications with increased safety requirements, it is also possible to design a redundant system with two isolated signal paths and two ASICs. The sensor principle can also be configured very advantageously in an “open” form as a travel sensor. The sensor can therefore be used at many points in a motor vehicle (e.g. recording the throttle-valve angle in the throttle device, lever position on auto-

Figure 7: Protective circuit of sensor with rotating alternating fields



matic transmissions, headlight position for headlight leveling control).

Magnetostatic sensors: Overview
Magnetostatic sensors measure a DC magnetic field. In contrast to the magnetically inductive sensors with coils, they are far more suitable for miniaturization and can be manufactured at reasonable cost using microsystem techniques. Since DC magnetic fields can easily penetrate through housing walls made of plastic but also of non-ferromagnetic metal, magnetostatic sensors have the advantage that the sensitive, generally fixed part can be well encapsulated and protected in respect of the rotatory part – usually a permanent magnet or a soft-magnetic conductive element – and in respect of the environment. Above all, galvanomagnetic effects (Hall and Gaussian effects) and magnetoresistive effects (AMR and GMR) are used.

Hall effect
The Hall effect occurs in thin semiconductor chips. If such a current-carrying chip is permeated vertically by a magnetic induction B, the charge carriers are deflected by the Lorentz force vertically to the field and to the current I by the angle α from their otherwise straight path. It is thus possible to pick off a voltage U_H proportional to the field B and to the current I transversally to the current direction between two opposing marginal points of the chip (Hall voltage, Figure 8):

$$U_H = \frac{R_H I B}{d}$$

where
R_H Hall coefficient,
d Chip thickness.

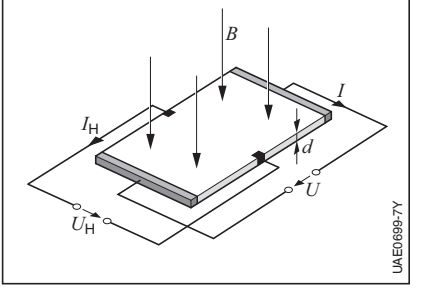
The coefficient R_H crucial to the measurement sensitivity of the chip is only comparatively small in the case of silicon. However, because the chip thickness d can be made extremely thin using diffusion techniques, the Hall voltage U_H attains a technically acceptable value again.

When silicon is used as the basic material, a signal-conditioning circuit can be integrated on the chip. Sensors using these principles are thus cheap to produce.

In terms of measurement sensitivity, temperature coefficient, and temperature range, however, silicon is by no means the best semiconductor material for Hall sensors. III-V semiconductors such as gallium arsenide or indium antimonide, for example, demonstrate better properties.

Gaussian effect
In addition to the transversally directed Hall effect, semiconductor chips are also subject to a longitudinal resistance effect, otherwise known as the Gaussian effect. Independently of the field direction the series resistance increases according to a roughly parabolic characteristic curve. Elements which utilize this effect are called magnetoresistors (Siemens trade name) and are manufactured from a III-V semiconductor, crystalline indium antimonide (InSb).

Figure 8: Hall effect
B Magnetic flux density,
I Supply current,
U Supply voltage,
U_H Hall voltage,
I_H Hall current.



Hall sensors

Hall switch

In the most simple case, in the Hall switch the Hall voltage is applied to an electronic threshold circuit integrated in the sensor (Schmitt trigger), which then outputs a digital signal. If the magnetic induction B applied at the sensor is below a given minimum threshold level, the Schmitt trigger's output value corresponds to a logical 0 (release status); if it exceeds a given upper threshold, the output value corresponds to a logical 1 (operate status). Since this behavior is guaranteed across the complete operating-temperature range and for all sensors of a given type, the two threshold values are relatively far apart (approx. 50 mT). In other words, it takes a considerable induction jump ΔB to trigger the Hall switch.

Such Hall sensors manufactured for the most part in bipolar technology are indeed very economical, but are at best only good for switch operation (e.g. Hall vane switches for triggering ignition in earlier ignition systems, digital steering-angle sensor). Hall vane switches are too inaccurate for recording analog variables.

Hall sensors using the spinning-current principle

A disadvantage of a simple Hall sensor is its simultaneous sensitivity to mechanical stresses (piezoelectric effect), which are inevitable due to packaging and result in an unfavorable offset temperature coefficient. This disadvantage has been overcome by the application of the spinning-current principle (Figure 9), combined with the transition to CMOS technology. The piezoelectric effect does arise here too, but is compensated by time averaging of the signal, since it occurs in the event of very fast, electronically controlled switching (rotation) of the electrodes with a different polarity sign.

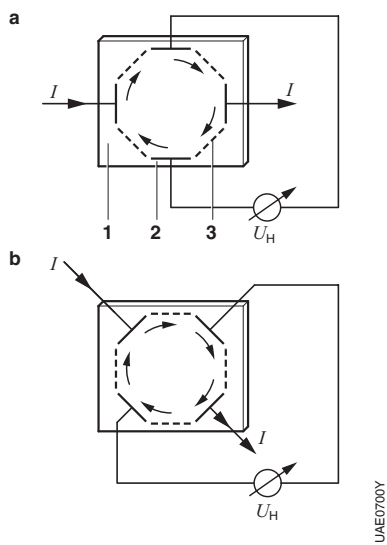
If the desired aim is to avoid the expense of complex electronics to switch the electrodes, it is also possible to integrate in close proximity several Hall sensors (two, four or eight) with correspondingly different alignment of the current path, and to add up their signals in the form of an averaging.

Only through these measures was it possible for Hall ICs to become well suited to analog sensor applications. These measures though barely resulted in a reduction of the sometimes considerable effects of temperature on the measurement sensitivity.

Such Hall ICs are suitable above all for the measurement of small travel distances, in which they register the fluctuating field strength of a permanent magnet as it approaches (used for example in the force sensor for sensing front-passenger weight). Up to then, similarly good results could only be achieved by the use of individual Hall elements of, for example, III-V semiconductors (e.g. GaAs) with a hybrid downstream amplifier.

Figure 9: Hall sensor using the spinning-current principle

a) Rotation phase φ_1,
b) Rotation phase φ_2, = φ_1 + 45°.
1 Semiconductor chip,
2 Active electrode,
3 Passive electrode.
I Supply current,
U_H Hall voltage.



Applications of Hall sensors
Position sensors
for transmission control
In this type of position sensor for transmission control for linear position sensing four Hall switches are arranged on a printed-circuit board in such a way that they sense the magnetic encoding of a linearly displaceable multipolar permanent magnet (Figure 10). The magnet carriage is coupled to the linearly actuated selector valve (hydraulic valve in the transmission control plate) or the parking-lock cylinder.

The position sensor senses the positions of the selector valve (P, R, N, D, 4, 3, 2) and the intermediate ranges and outputs these in the form of a 4-bit code to the transmission-control system. For safety reasons the encoding of the position is configured as a single step. In other words, two bit changes are always required until a new position is detected. The sequence of these bit changes corresponds to a Gray code.

Axle sensors
Axle sensors sense the body's inclination angle that changes as a result of vehicle load or during braking and acceleration. The range of the headlamps can be adapted to the conditions with this information (automatic headlamp leveling control).

Vehicle inclination is measured by angle-of-rotation sensors (axle sensors). These are mounted to the body at the front and rear. Compression is measured by a rotary lever which is connected via a pushrod to the vehicle axle. The vehicle inclination is obtained from the difference between the sensor signals on the front and rear axles.

The axle-sensor stator (Figure 11) incorporates a Hall element located in the ring magnet's homogeneous field. The magnetic field induces in the Hall element a Hall voltage that is proportional to the magnetic field strength. During compression the axle-sensor shaft rotates with the ring magnet, and thus the magnetic field is also altered by the Hall element. The Hall element delivers a signal that is dependent on the angle of rotation of the shaft and thus on the compression travel.

Accelerator-pedal sensors
Using a rotatable magnetic ring ("movable magnet"), together with a number of fixed

Figure 10: Encoding of position sensor for transmission control
a) Magnetic encoding,
b) Position ranges.
1 Moving carriage,
2 Fixed position of Hall elements.

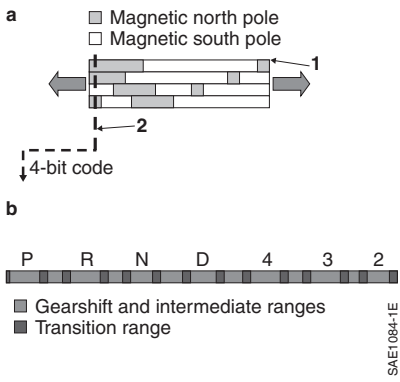
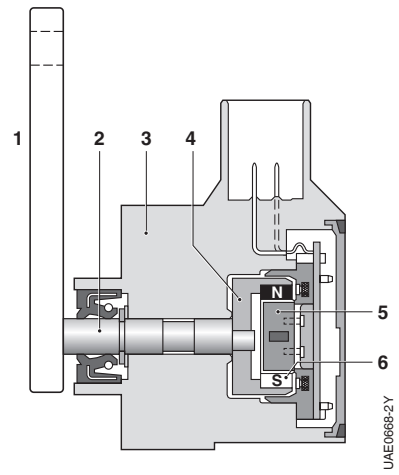


Figure 11: Axle sensor
1 Rotary lever,
2 Shaft,
3 Housing,
4 Ring-magnet mounting,
5 Stator with Hall IC,
6 Ring magnet.



soft-magnetic conductive elements, a linear output signal can be generated for a larger angular range without conversion being necessary (Figure 12). Here, the movable magnet's bipolar field is directed through a Hall sensor located between semicircular conductive elements. The effective magnetic flux through the Hall-effect sensor is dependent on the angle of rotation φ. The angular range that can be sensed is 180° here.

The Hall angular-position sensor with a measuring range of approximately 90° pictured in Figure 13a is derived from the basic "movable magnet" principle. The magnetic flux from a practically semicircular permanent-magnet disc is returned to the magnet through a pole-shoe, two additional conductive elements each of which contains a Hall sensor in its magnetic path, and the shaft which is also ferromagnetic. Depending upon the angular setting, the flux is led to a greater or lesser degree through the two conductive elements, in the magnetic path of which a Hall sensor is situated. Using this principle, it is possible to achieve a practically linear characteristic curve (Figure 13b).

The simplified version pictured in Figure 14 does without soft magnetic conductive elements. In this version, the magnet moves around the Hall sensor in a circular arc (Figure 14a). Only a relatively small section of the resulting sinusoidal characteristic curve features good linearity. If the Hall sensor is located slightly outside the center of the circular arc, the characteristic curve increasingly deviates from the sinusoidal, and now features a short measuring range of almost 90°, and a longer measuring range of something more than 180° with good linearity. A great disadvantage though is the low level of shielding against external fields, as well as the remaining dependence on the geometric tolerances of the

Figure 12: Angular-position sensor using the movable-magnet principle
a) Sensor in initial position,
b) Excursion through angle of rotation φ,
c) Output signal.
1 Magnetic yoke,
2 Stator (soft iron),
3 Rotor (permanent magnet),
4 Air gap,
5 Hall sensor.
φ Angle of rotation.

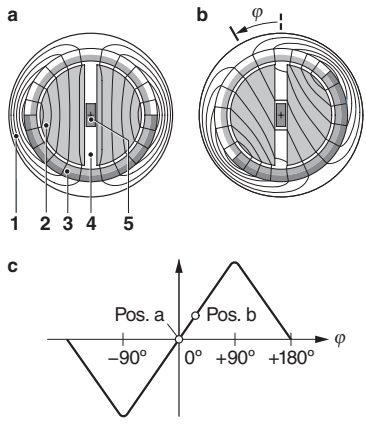
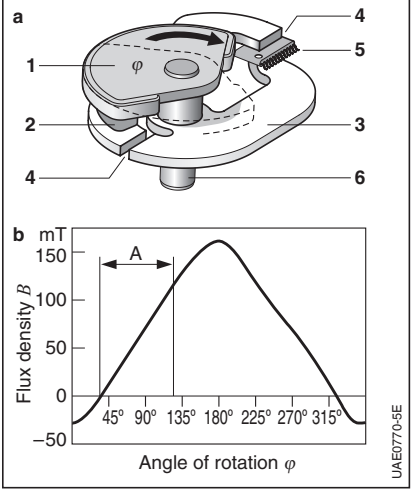


Figure 13: Hall angular-position sensor with linear characteristic curve up to 90° using the movable-magnet principle
a) Layout,
b) Characteristic curve with working range A.
1 Rotor disk (permanent magnet),
2 Pole shoe,
3 Conductive element,
4 Air gap,
5 Hall sensor, 6 Shaft.
φ Angle of rotation.



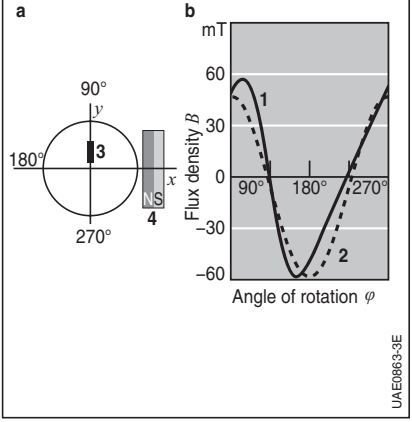
magnetic circuit, and the intensity fluctuations of the magnetic flux in the permanent magnet as a function of temperature and age.

In the Hall angular-position sensor pictured in Figure 15 it is not the field strength but rather the magnetic-field direction that is evaluated. The magnetic-field lines are sensed by four Hall elements lying in one plane and radially arranged in the x- and y-directions. The sensor is located between two magnets to generate a homogeneous magnetic field. The sensor signal is formed from the measurement signals of the Hall elements (sine and cosine signals) in an evaluation circuit using the arctan function.

Like the potentiometric accelerator-pedal sensor these systems also contain two measuring elements in order to obtain two redundant voltage signals for diagnostic purposes.

Figure 14: Hall angular-position sensor with linear characteristic curve up to 180° using the movable-magnet principle

a) Principle (Hall IC in this depiction shifted from center point),
b) Characteristic curve.
1 Curve for arrangement with Hall IC at center point,
2 Curve for arrangement with Hall IC outside center point,
3 Hall IC,
4 Magnet.



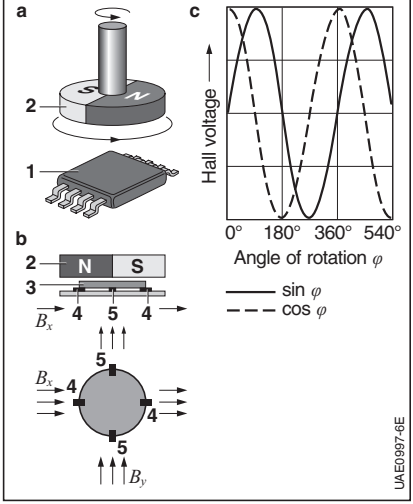
Magnetoresistive sensors

In contrast to the Hall sensors, the optimum chip shape for a magnetoresistor tends to be shorter and squatter, and represents a very low resistance. In order to arrive at technically applicable resistance values in the kΩ range, it is therefore necessary to connect a large number of these chips in series. This problem is solved elegantly by adding microscopically fine, highly conductive nickel-antimonide needles to the semiconductor crystal. These are located obliquely to the direction of current flow. A further measure is to apply meander techniques to the semiconductor resistor.

The dependence of the resistance on the magnetic flux density B follows a square-law function up to inductances of approximately 0.3 T, and above this point

Figure 15: Hall angular-position sensor with fourfold Hall sensor with evaluation of magnetic-field direction for a measurement range over 360°

a) Layout,
b) Measurement principle,
c) Measurement signals.
1 IC with Hall elements,
2 Magnet (the opposite magnet is not visible here),
3 Conductive element,
4 Hall elements for sensing the x component of B,
5 Hall elements for sensing the y component of B.



it is increasingly linear. There is no upper limit to the control range, and dynamic response in technical applications can – as is the case with the Hall sensor – be regarded as free from lag.

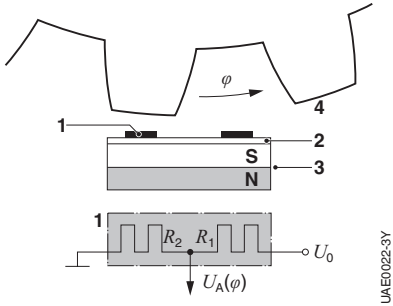
Since the temperature sensitivity has a pronounced effect on the resistance of magnetoresistors (approximately 50 % reduction for 100 K), they are usually delivered only in the dual-configuration form in voltage-divider circuits (differential magnetoresistors). For the particular application, each of the two resistor sections must then be magnetically triggered (as far as possible with opposite polarities). Notwithstanding the high temperature coefficient of the individual resistors, the voltage-divider circuit guarantees good stability of the working-point (that point at which both resistor sections have the same value).

In order to achieve good measurement sensitivity, it is best to operate the magnetoresistors at a magnetic working point between 0.1 and 0.3 T. Generally, the required magnetic bias is supplied by a small permanent magnet (Figure 16) the effects of which can be increased by using a small magnetic return plate. The magnetoresistor features pronounced temperature sensitivity so that it is used almost exclusively in incremental angle-of-rotation and displacement sensors, or in binary limit-value sensors (with switching characteristic). A magnetoresistive increment sensor is used to measure start-of-injection adjustment in diesel distributor injection pumps.

The magnetoresistor's main advantage is its high signal level which is usually in the volts range. This means that amplification is unnecessary, as well as the local electronic circuitry and the associated protective measures which would otherwise be needed. Furthermore, in their role as passive, resistive components, they are highly insensitive to electromagnetic interference and, as a result of their high bias voltage, practically immune against external magnetic fields.

Figure 16: Differential magnetoresistive sensor

1 Magnetoresistors R_1 and R_2,
2 Soft-magnetic substrate,
3 Permanent magnet, 4 Gear.
U_0 Supply voltage,
$U_A(\varphi)$ Measurement voltage for angle of rotation φ.



AMR sensors

Anisotropic magnetoresistive effect

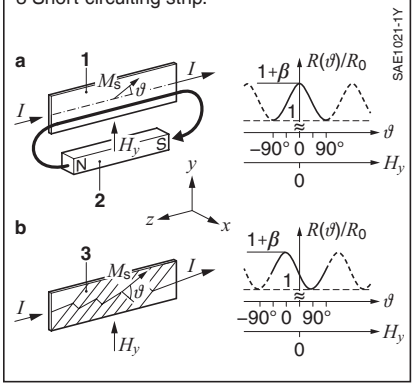
NiFe layers only roughly 30 to 50 nm thick display electromagnetically anisotropic properties, i.e. their electrical resistance changes under the influence of a magnetic field. Resistance structures of this type are therefore called AMR sensors (anisotropic magnetoresistive sensors). The metal alloy generally used is also called permalloy.

Types

On an elongated resistance strip, as shown in Figure 17a, a small, spontaneous magnetization M_S occurs in the longi-

Figure 17: AMR basic principle

a) Basic type,
b) Barber-pole type.
1 Magnetoresistive element (NiFe, NiCo),
2 Bias magnet,
3 Short-circuiting strip.



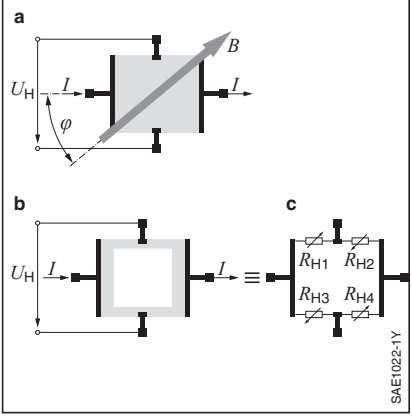
tudinal direction of the conductor track even without the application of an external control field (form anisotropy). In order to give this magnetization a clearly defined direction – theoretically, it could be in the other direction – AMR sensors are often, as depicted, provided with weak bias magnets. In this state, the series resistance has its greatest value $R_{\|}$. If the magnetization vector is rotated under the influence of an additional external field H_y through the angle ϑ, the series resistance decreases gradually until it assumes at $\vartheta = 90°$ a minimum value $R_{\perp}$. Here, the resistance is only dependent on the angle ϑ contained by the resulting magnetization M_S and the current I; it has a roughly cosinusoidal characteristic as a function of ϑ:

$$R = R_0(1 + \beta\cos^2\vartheta).$$

Thus the following relationships are obtained for the maximum and minimum values:

$$R_{\|} = R_0(1 + \beta) \text{ and } R_{\perp} = R_0.$$

Figure 18: Pseudo-Hall sensor
a) Full basic form,
b) Modified form with hollowed-out surface,
c) Replacement circuit diagram for b).
B Magnetic flux density,
U_H Hall voltage,
I Supply current,
φ Angle of rotation,
R_H Resistors of AMR elements.



The coefficient β denotes the maximum possible resistance variation, and is approximately 3 %. If the external field is much stronger than the spontaneous magnetization, and this is usually the case when control magnets are used, the effective angle is almost completely a function of the direction of the external field. The field strength is now irrelevant, in other words, the sensor is now operating in the "saturated state".

Highly conductive short-circuiting strips (for instance of gold) on the AMR film force the current to flow at an angle of below 45° to the spontaneous magnetization (longitudinal direction) without the application of an external field. In this version – also referred to as the "barber-pole" sensor – the sensor curve shifts by 45° compared to that of the simple resistor (Figure 17b). This means, therefore, that even when the external field strength $H_y = 0$, the curve is at the point of maximum measurement sensitivity (reversal point). The "striping of two resistors in opposite directions" means that they change their resistances in opposite directions under the effects of the same field. In other words, one of the resistances increases while the other drops.

In addition to the simple, two-pole AMR elements, there are also pseudo-Hall sensors, for instance square NiFe thin-film structures, which, like the normal Hall sensors already described, have four terminals: two for the current path and two transverse ones for picking off a Hall voltage (Figure 18a). Unlike a normal Hall sensor, however, a pseudo-Hall sensor is sensitive to magnetic fields in the film plane and not vertically to it. Also, a pseudo-Hall sensor does not demonstrate a proportional characteristic curve; instead, it has a sinusoidal curve with very high faithfulness of shape to the sine, which is in no way dependent on the strength of the control field and the temperature. For a field parallel to the current path, the output voltage disappears in order then to describe a sinusoidal half-period when the field is rotated up to the angle $\varphi = 90°$. The sinusoidal voltage therefore results with amplitude $\hat{u}_H$ as follows:

$$U_H = \hat{u}_H \sin 2\varphi.$$

If the external control field is rotated once through φ = 360°, the output voltage therefore follows two full sinusoidal periods. The amplitude $\hat{u}_H$ is, however, very much dependent on the temperature and the air gap between the sensor and the control magnet; it decreases as the temperature and the air gap increase.

The measurement sensitivity of this so-called pseudo-Hall element can be considerably increased (without excessively falsifying the sinusoidal shape) by "hollowing out" the element from the inside so that only the "frame" remains (Figure 18b). This modification converts the pseudo-Hall sensor to a full bridge consisting of four AMR resistors (Figure 18c). Even when the bridge resistors are meander-shaped, provided a given minimum conductor width is not dropped below, this still has negligible effect on the signal's sinusoidal shape.

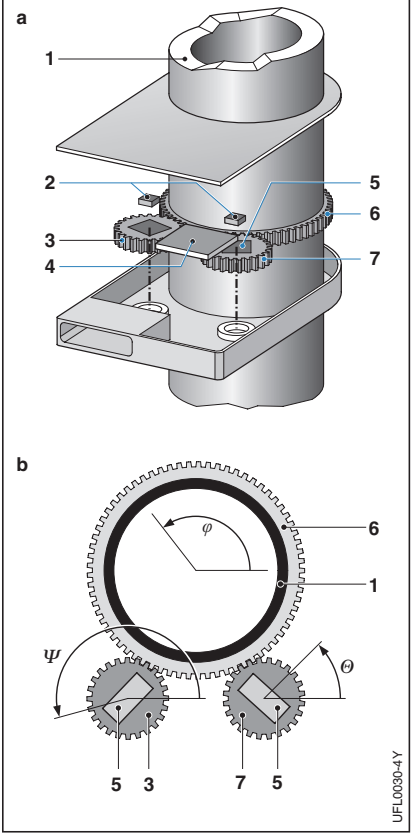
Applications

Steering-angle sensor

The function of driving-dynamics systems is to keep the vehicle on the desired course specified by the driver with targeted brake interventions. The steering-wheel angle must be known for this purpose. Potentiometers, optical code sensing and magnetic measurement principles are used for measurement.

A steering-wheel-angle sensor mounted on the steering shaft senses the steering-wheel position. Several rotations of the steering shaft can be measured with a dual configuration of pseudo-Hall angle-of-rotation sensors (each for 180 °). The two associated permanent magnets are rotated via a high-ratio gear train (Figure 19). Because the two smaller output gears, which carry the control magnets, differ by one tooth, their mutual phase angle (difference of angles of rotation: $\Psi - \Theta$) is a clear measure of the absolute angular position of the steering shaft. The system is designed in such a way that this phase difference does not exceed 360° for a total of four rotations of the steering shaft, and thereby the clearness of the measurement is assured. Each individual sensor also offers an indeterminate fine resolution of the angle of rotation. It is possible with such an arrangement to resolve the full steering-angle range, for example, more accurately than 1°.

Figure 19: AMR steering-angle sensor
a) Layout,
b) Angle relations.
1 Steering shaft,
2 AMR measuring cells,
3 Gear with m teeth,
4 Evaluation electronics,
5 Magnets,
6 Gear with $n > m$ teeth,
7 Gear with $m + 1$ teeth.



GMR sensors

Giant magnetoresistive effect

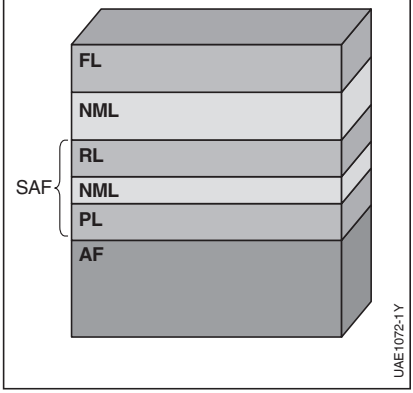
GMR-sensor technology (giant magnetoresistive sensor) is used in automotive applications for angular-position and rpm sensing. The main advantages of GMR sensors over AMR sensors are the natural clearness range within 360° for angular-position sensing and the higher magnetic-field sensitivity for rpm sensing.

GMR layer structures consist of antiferromagnetic, ferromagnetic and non-magnetic function layers (Figure 20). The individual layer thicknesses in both systems are in the range of 1 to 5 nm, and therefore only comprise a few atom layers. For angular-position sensing the required reference magnetization is generated by the fact that the direction of magnetization of one of the ferromagnetic layers (PL) is pinned by the interaction with a neighboring antiferromagnetic layer (AF). This is therefore also referred to as a “pinned layer”. On the other hand, the magnetization of the second ferromagnetic layer (FL) magnetically isolated to a large extent via a non-magnetic interlayer (NML) can be freely rotated with the external magnetic field. This is accordingly referred to as a “free layer”.

The resistance changes with a cosinusoidal dependence on the angle between the external field direction and the reference direction. Crucial to the accuracy of the angular-position measurement is the stability of the reference magnetization against the effect of the external field. This stability is greatly increased by using an additional artificial antiferromagnet (SAF).

Figure 20: GMR layer structure

FL Free layer,
NML Non-magnetic interlayer,
RL Reference layer,
PL Pinned layer,
AF Antiferromagnet,
SAF Artificial antiferromagnet.



Applications of GMR sensors

Steering-wheel-angle sensor

The mechanical structure and the operating principle of this GMR sensor is equivalent to those of the steering-wheel-angle sensor with AMR elements. In view of the greater sensitivity in comparison with the AMR effect the GMR sensor can operate with weaker magnets and larger air gaps. This results in costs benefits with regard to materials and design. The 360° angular-position measurement range of a single GMR element (180° is typical for AMR) enables smaller gears to be used. This in turn means that it takes up less space.

References for position sensors

[1] K. Reif (Editor): Klassische Diesel-Einspritzsysteme – Bosch Fachinformation Automobil. 1st Ed., Vieweg+Teubner, 2012.
[2] K. Reif (Editor): Sensoren im Kraftfahrzeug – Bosch Fachinformation Automobil. 3rd Ed., Verlag Springer Vieweg, 2016.

rpm sensors

Measured variables
rpm sensors measure the time required for an angle to be covered during a rotational motion. From this the number of revolutions per unit of time can be determined. In a motor vehicle this usually involves relative measured variables which occur between two components. Examples include:
– Crankshaft speed,
– Camshaft speed,
– Wheel speed (e.g. for the antilock braking system),
– Transmission speed,
– Rotational speed of the diesel distributor-type injection pump.

Measurement principles
rpm sensing usually occurs with an incremental sensor system, consisting of a rotor (e.g. gear or multipole wheel) and the rpm sensor (Figure 1).

Passive rpm sensors
Conventional inductive sensors used in the past were based on the inductive measuring effect. A sensor of this type consists of a permanent magnet and a soft magnetic pole pin surrounded by an induction coil. The pole pin is opposite a ferromagnetic gear. The distance between the gear and the pole pin changes as the gear turns. The time change in the magnetic flux that this causes gives rise to an induced voltage in the coil.

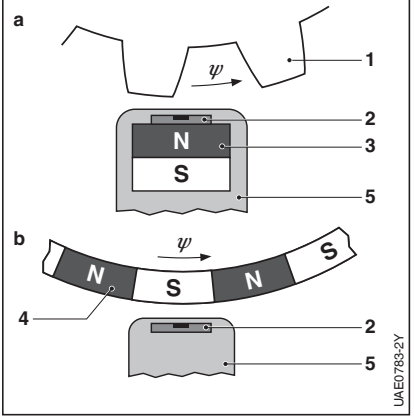
The measuring effect of the inductive sensor is relatively high, eliminating the need for local electronics. It is therefore called a passive sensor. However, the signal amplitude is rotational-speed-dependent. The sensor is therefore unsuitable for the lowest rotational speeds; it permits only a comparatively small air-gap tolerance between pole pin and gear and for the most part is not able to distinguish air-gap fluctuations (chatter) from rotational-speed pulses.

Passive rpm sensors are still used today in commercial vehicles.

Active rpm sensors
Active rpm sensors operate in accordance with the magnetostatic principle. The amplitude of the output signal is not dependent on the rotational speed. rpm sensing is thus possible even at very low rotational speeds (quasistatic rpm sensing).

Hall sensor
A voltage U_H proportional to the magnetic field (Hall voltage) can be picked off transversely to the current direction on a current-carrying wafer which is permeated vertically by a magnetic induction B (see Hall effect). In the sensor arrangement with a ferromagnetic gear (pulse wheel) the magnetic field is generated by a permanent magnet (Figure 1a). Located between the magnet and the pulse wheel is the Hall-sensor element. The magnetic flux which permeates this element depends on whether the sensor is opposite a tooth or a gap. This results in a Hall voltage that is equivalent to the progression of the teeth. This rotational-speed infor-

Figure 1: Hall sensors
a) Sensor arrangement with passive rotor (ferromagnetic gear),
b) Sensor arrangement with active rotor (multipole wheel).
1 Increment rotor,
2 Hall IC,
3 Permanent magnet,
4 Multipole wheel,
5 Housing.
ψ Rotational speed.



mation is conditioned and amplified, and then transmitted as injected current in the form of square-wave signals. Typical values are 7 mA for the low level and 14 mA for the high level. The current is converted into a signal voltage by a measuring shunt in the ECU.

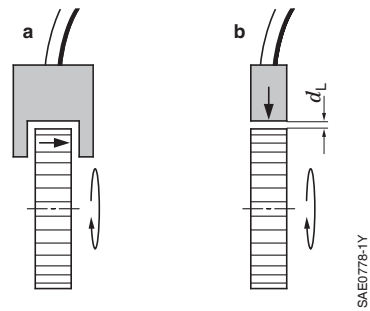
In a differential Hall sensor there are two Hall-sensor elements situated between the magnet and the pulse wheel. Determining the difference between the signals from the two sensor elements reduces magnetic interference signals and improves the signal-to-noise ratio.

Multipole wheels are also used instead of the ferromagnetic pulse wheel. Here, a magnetizable plastic is applied to a non-magnetically metallic substrate and alternately magnetized. These north and south poles assume the function of the teeth on a pulse wheel (Figure 1b).

AMR sensor

The electrical resistance of magnetoresistive material (AMR, Anisotropic Magneto Resistive) is anisotropic, i.e. it is dependent on the direction of the magnetic field to which it is exposed. This property is exploited in the AMR sensor. The sensor is located between a magnet and the pulse wheel. The magnetic field lines change their direction when the pulse wheel turns. This gives rise to a sinusoidal voltage which is amplified in an evaluation circuit in the sensor and converted into a square-wave signal.

Figure 2: Different sensor shapes
a) Fork shape (vane principle),
b) Rod shape (proximity principle).
d_L Air gap.



GMR sensor

The use of GMR technology (Giant Magneto Resistance) constitutes a further development of active sensors. Their greater sensitivity than AMR sensors means that larger air gaps are possible, which in turn opens up the possibility of use in difficult areas of application. The greater sensitivity also means lower noise at the signal edges.

Sensor shapes

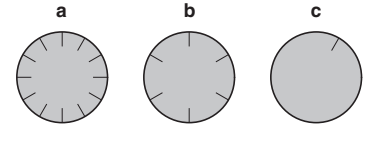
Different sensor shapes are used (Figure 2): the rod shape, the fork shape, and the inner and outer ring shape. Due to its ease of mounting and simplicity, the rod sensor is the most widespread sensor shape. The rod sensor is located near the rotor, the teeth of which approach it and pass by in close proximity. However, it also has the lowest measurement sensitivity. The fork-shaped sensor, which is insensitive to axial and radial play, is permissible in some cases, and is also in service in the field. This sensor must be roughly aligned to the rotor when installed. The sensor type in which the sensor surrounds the rotor shaft in the form of a ring is practically no longer used.

Rotor shapes

In the incremental detection of relative yaw rate, depending upon the number and size of the scanned peripheral rotor markings, a distinction is made between the following types of sensor (Figure 3):

– Simple sensor, with only a single scanned marking per revolution, so that only the average rotational speed can be registered,

Figure 3: Recording of relative rotational speed
a) Increment sensor,
b) Segment sensor,
c) Simple rpm sensor.



– Segment sensor, with only a small number of scanned peripheral segments (for instance, equivalent to the number of engine cylinders),
– Increment sensor with closely spaced peripheral markings. Up to a certain point, this form of sensor permits instantaneous speed to be measured at points on the circumference, and thus the registration of very fine angular divisions.

The rotor is crucially important to measuring rotational speed; however, it is usually part of the vehicle manufacturer's delivery specification, while the actual sensor comes from the supplier. The rotors used almost exclusively up to now have been magnetically passive types, which therefore consist of soft-magnetic material, usually iron. They are cheaper than hard magnetic pole wheels and are easier to handle, since they are not magnetized, and there is no danger of mutual demagnetization during storage. As a rule, presuming the same increment width and output signal, the pole wheel's intrinsic magnetism (a pole wheel is defined as a magnetically active rotor) permits a considerably larger air gap.

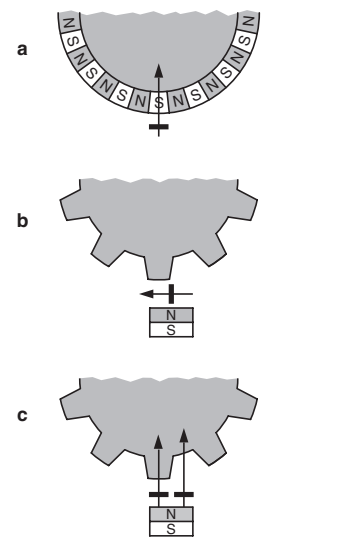
Demands made of rpm sensors

The following demands are made of rpm sensors:
– Static detection (i.e. speed approaching zero, extremely low engine-starting and wheel speeds),
– Large air gaps (non-adjusted mounting on air gap greater than zero),
– Small size,
– Efficient operation independent of air-gap fluctuations,
– Heat-resistant up to 200 °C,
– Identification of the direction of rotation (optional for navigation),
– Reference-mark identification (ignition).

Magnetostatic sensors (e.g. Hall sensors, AMR sensors) are suitable for satisfying the first condition. And, as a rule, they also permit compliance with the second and third stipulations.

Figure 4 shows three basically suitable sensor shapes which generally are insensitive to air-gap fluctuations. Here, a distinction must be made between sensors which sense radially and those which sense tangentially (see also Figure 5). This means that independent of the air gap, magnetostatic sensors are always able to differentiate between the north and south poles of a magnetically active pole wheel. In the case of magnetically passive rotors, the sign of the output signal is then no longer independent of the air gap when they register the tangential-field strength (here though, the fact that the air gap is often enlarged due to the rotor is a disadvantage). Here, though, the fact that the air gap between the rotor and the permanent magnet is often enlarged due to the sensor is a disadvantage. However, types frequently used as well are radially measuring differential-field or gradient sensors, which essentially only register the gradient of the radial field component,

Figure 4: Sensor arrangements which are insensitive to air-gap fluctuations
a) Radial-field sensor with pole wheel,
b) Tangential sensor,
c) Differential sensor with gear.
The sensor measures that part of the magnetic field in a certain direction in space. The arrow marks this direction.



which changes with regard to its sign not with the air gap, but only with the angle of rotation.

In the case of wheel-speed sensors, at the least the sensor tip, because of its proximity to the brake, should be able to withstand higher temperatures.

Sensor arrangements

Gradient sensors

Gradient sensors (e.g. based on differential Hall sensors) incorporate a permanent magnet on which the pole surface facing the gear is homogenized with a thin ferromagnetic wafer. Two sensor elements are located on each element's sensor tip, at a distance of roughly one half a tooth interval. Thus, one of the elements is always opposite a gap between teeth when the other is opposite to a tooth. The sensor measures the difference in field intensity at two adjacent locations on the circumference. The output signal is roughly proportional to the diversion of field strength as a function of the angle at the circumference; polarity is therefore independent of the air gap.

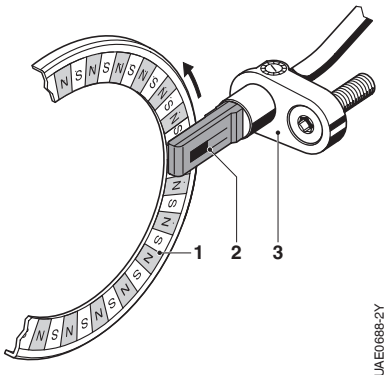
Tangential sensors

Tangential sensors differ from gradient sensors by their reaction to variations in polarity and intensity in the components of a magnetic field located tangentially to the periphery of the rotor. Design options include AMR thin-film technology (barber pole) or single permalloy resistors featuring full- or half-bridge circuits [1]. In contrast to gradient sensors, they need not be matched to the particular tooth pitch of the rotor and can be designed to sense at a given point. Local amplification is necessary, even though their measuring effect is 1 to 2 orders of magnitude larger than that of the silicon-Hall sensors.

In the case of a bearing-integrated crankshaft speed sensor (Simmer shaft-seal module), the AMR thin-film sensor is mounted together with an evaluation IC on a common lead frame. For the purposes of space saving and temperature protection, the evaluation IC is bent at an angle of 90° and also located further away from the sensor tip.

Figure 5: AMR rpm sensor as tangential-field sensor

1 Multipole ring,
2 Measuring cell,
3 Sensor housing.



Applications

Crankshaft-speed sensor

Crankshaft-speed sensors are used to measure the engine speed (crankshaft speed). In addition, the engine-management system requires the position of the crankshaft (position of the pistons) to be able for example to activate ignition coils and fuel injectors at the correct crank angle. The pulse wheel mounted on the crankshaft exhibits a tooth gap for this purpose. With a tooth pitch of 6° and two missing teeth the pulse wheel has 58 teeth. The rpm sensor delivers a square-wave signal (tooth signal) that corresponds to the teeth on the pulse wheel.

The microcontroller in the engine ECU records the value of a running timer at the falling edges of the signal and forms the tooth spacing in terms of time from the difference between two edges. At the first tooth after the gap the measured time is very much greater than the preceding time; at the second tooth after the gap the time is again very much less. This tooth is defined as the reference mark. The pulse wheel is mounted so that this position corresponds to a defined angle opposite the top dead center of cylinder no. 1.

At uniform rotational speed the time between the edges around the tooth gap is three times as long as between two normal edges. Even in the event of greatly fluctuating rotational speed when

the engine is being started, the reference mark can be reliably inferred on account of the measured differences of the tooth spacing.

Camshaft-speed sensor
The camshaft is stepped down by a ratio of 1:2 compared with the crankshaft. It indicates whether a piston moving to top dead center is in the compression cycle or the exhaust cycle. For this reason a camshaft-speed sensor is required in addition to the crankshaft-speed sensor.

Simple pulse wheels lead to different sensor-signal levels only in the area of the reference mark. More complex pulse wheels with segments of different lengths enable the engine position to be detected more quickly when starting and thus provides for rapid starting.

Wheel-speed sensor
Wheel-speed sensors sense the rotational speed of the wheels. This information is needed for example by the driving-dynamics control to calculate the driving speed and the slip. The navigation system requires the wheel-speed signals to calculate the distance traveled when there is no GPS reception (e.g., in tunnels).

The pulse wheel is permanently connected to the wheel hub. Both steel pulse wheels and multipole rings can be used. This sensor has fewer teeth because the pulse wheel is smaller in diameter than that on an engine rpm sensor.

The sensor element with the signal amplifier and signal conditioning is integrated in an IC, hermetically sealed with plastic, and seated in the sensor head.

Digital signal conditioning enables encoded additional information to be transmitted via a pulse-width-modulated output signal. One example is direction-of-rotation detection, which is required for Hill Hold Control that prevents the vehicle from rolling back when starting. The direction of rotation is also used for vehicle navigation to detect backing up (reversing).

Speed sensor for transmission control
Transmission-speed sensors sense shaft speeds in automatic transmissions. Standard geometries are generally not possible due to the compact designs of transmissions. Specific sensor versions are therefore required for each transmission. Evaluation circuits with varying degrees of evaluation-algorithm complexity are used to cover the full spectrum of functional requirements.

References for rpm sensors
[1] K. Reif (Editor): Sensoren im Kraftfahrzeug. 2nd Ed., Springer-Vieweg, 2012.

Oscillation gyrometers

Measured variables

Oscillation gyrometers measure the absolute yaw rate along the vehicle vertical axis, the vehicle transverse axis (pitch axis), and the vehicle longitudinal axis (roll rate). The signals generated here are used in numerous systems such as for example driving-dynamics control, rollover protection systems, vehicle navigation, and damping control.

Measurement principle

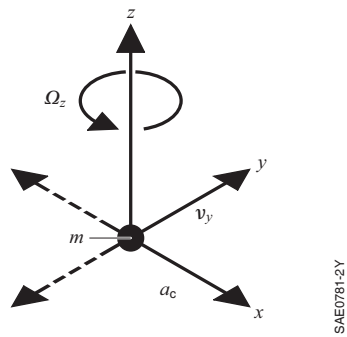
Oscillation gyrometers are similar in principle to mechanical gyroscopes and for measurement purposes utilize the Coriolis acceleration a_c that occurs during rotary motion in conjunction with an impressed oscillating motion (Figure 1). The acceleration is calculated from the cross product of velocity and yaw rate:

(1) $\vec{a}_c = \vec{a}_x = 2\vec{v}_y \times \vec{\Omega}_z$.

The velocity v is produced by the sinusoidal driving of the inert mass of the sensor along its y-axis at frequency ω. Thus:

(2) $v_y = \hat{v}_y \sin(\omega t)$.

Figure 1: Creation of Coriolis acceleration
When a mass point m moves in the y-direction at velocity v_y and the system simultaneously rotates about the vertical axis z with yaw rate Ω_z, the mass point experiences a Coriolis acceleration a_c in the x-direction.



With a constant yaw rate Ω about the z-axis a Coriolis acceleration is produced along the x-axis. It has the same frequency and phase as the drive velocity v, and the associated amplitude is

(3) $\hat{a}_c = 2\,\hat{v}_y \Omega_z$.

The Coriolis force is one of the so-called pseudo forces and can only be measured in the rotating reference system (like the passenger car here).

The signal evaluation of the sensor demodulates the sinusoidal Coriolis signal of the micromechanical sensor ($a_c \propto \sin(\omega t)$) and in this way determines the yaw rate Ω. In the process, unwanted acceleration from the outside (e.g. bodywork acceleration) is removed.

The utilized measuring effects not only are very small, but also require complex signal conditioning. Signal conditioning directly at the sensor is therefore necessary.

Applications

Micromechanical yaw-rate sensors

A seismic mass in the sensor must be induced to vibrate in order to generate a Coriolis force to measure a yaw rate. In a micromechanical yaw-rate sensor (Figure 2) this is achieved in a permanent-magnet field by utilizing the Lorentz force (electrodynamic drive).

This sensor is based on a formulation of bulk micromechanics: Two thick oscillating elements fashioned from the wafer by means of bulk micromechanics oscillate in push-pull mode at their resonant frequency (< 2 kHz), which is determined by their mass and their coupling-spring stiffness. Each of them is provided with a capacitive, surface-micromechanical acceleration sensor (SMM) which measures Coriolis acceleration.

For drive purposes a sinusoidally modulated current is passed through a simple printed conductor on the relevant oscillating element. This generates within the permanent-magnetic field operating perpendicularly to the chip surface a Lorentz force which starts the oscillating element moving.

If this sensor is rotated at yaw rate Ω about the vertical axis, the drive in the oscillation direction results in a Coriolis acceleration which is measured by the acceleration sensors.

The different physical natures of the drive and sensor systems prevent unwanted crosstalk between the two parts. In order to suppress external acceleration (common-mode signal), the two opposing sensor signals are subtracted from each other. Summation, however, can also be used to measure the external acceleration).

The high-precision micromechanical construction helps to suppress the effects of high oscillatory acceleration which is several factors of 10 higher than the low-level Coriolis acceleration (cross sensitivity far below 40 dB). The drive and measuring systems are mechanically and electrically isolated in rigorous terms here.

Surface-micromechanical yaw-rate sensors

Micromechanical yaw-rate sensors can also be manufactured entirely using SMM technology (surface micromechanics). Here the oscillating element is electrostatically driven by capacitors. In this case the isolation of the drive and detection systems is not longer strictly separated from each other since both systems are capacitive in design.

A centrally mounted rotary oscillator is electrostatically driven to oscillate using comb structures (Figure 3). When the sensor is subjected to a rotation Ω, the Coriolis force induces a simultaneous "out-of-plane" tilting motion. Its amplitude is proportional to the yaw rate and is detected capacitively by electrodes located under the rotary oscillator. To prevent this motion from being excessively damped, it is essential to operate the sensor in a vacuum.

The smaller chip size and the simpler manufacturing process do indeed reduce the cost of such a sensor, but the reduction in size also diminishes the already slight measuring effect and thus the attainable accuracy. It places higher demands on the electronics.

Figure 2: Micromechanical yaw-rate sensor with electrodynamic drive

1 Oscillation direction,
2 Oscillating element,
3 Coriolis acceleration sensor,
4 Retaining and guide spring,
5 Direction of Coriolis acceleration,
Ω Yaw rate,
v Oscillation velocity,
B Magnetic flux density.

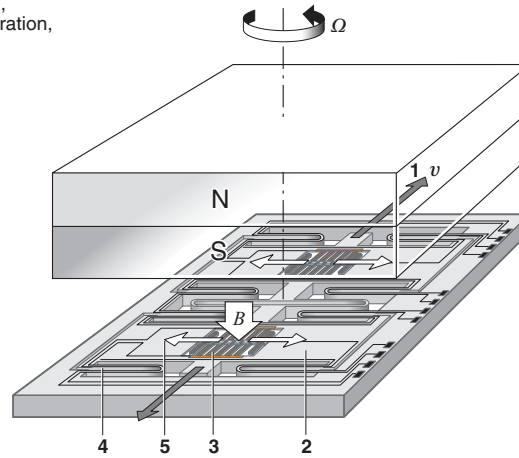
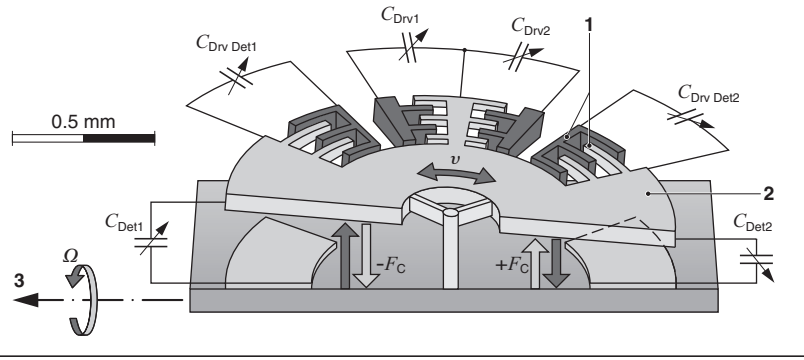


Figure 3: Surface-micromechanical yaw-rate sensor with electrostatic drive
1 Comb structure, 2 Rotary oscillator, 3 Measurement axis.
C_{Drv} Capacitance of drive electrodes (drive by application of a sinusoidal voltage),
C_{Det} Capacitance of rotary-oscillation tap (measurement of Coriolis force),
C_{DrvDet} Capacitance of drive tap (measurement of drive oscillation),
F_C Coriolis force,
v Oscillation velocity,
Ω Yaw rate to be measured (Ω = const · ΔC_{Det}).



The influence of external accelerations is suppressed in SMM yaw-rate sensors by a special micromechanical design. Here the MEMS element (MEMS, Micro-Electro-Mechanical Systems) exhibits two seismic masses. Their deflection is effected by Coriolis forces in opposite directions, but by external accelerations in the same direction. A differential evaluation of the two oscillators then frees the signal of the unwanted disturbance values.

Further developments of yaw-rate sensors

Ongoing development of systems in the field of active and passive safety means that increased demands are placed on the signal quality and robustness of the yaw-rate sensor. These systems require not only the measurement of the yaw rate, but also the detection of pitching and rolling motion.

For example, a driving-dynamics control system is based among other things on measurement of the yaw rate Ω_z and on measurement of the lateral acceleration a_y. ACC systems (Adaptive Cruise Control) utilize sensor signals of the yaw rate Ω_z, the lateral acceleration a_y, and the linear acceleration a_x. Rollover protection systems on the other hand use sensor information about the roll axis Ω_x, the vertical acceleration a_z, and also the lateral acceleration a_y.

New sensor generations of micromechanical sensor elements are used here. These sensors measure the yaw rate about either the x-axis or the z-axis; many of the these sensors also incorporate sensing elements for detecting accelerations in the x-, y- or z-direction.

Measurement of the yaw rate is based here on the principle of the surface-micromechanical yaw-rate sensor, while accelerations are detected with capacitive surface-micromechanical acceleration sensors.

Flow meters

Measured variables

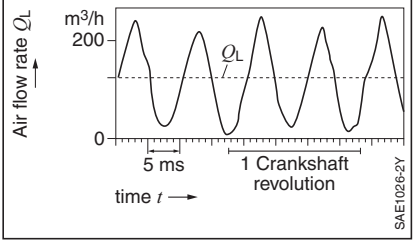
The amount of air supplied to the engine must be measured in order to control the combustion process. The mass ratios are the main factors in the chemical process of combustion, thus the actual objective is to measure the mass flow of the intake and charge air. The air-mass flow is the most important load variable in gasoline engines; precise pilot control of the air/fuel ratio presupposes knowledge of the supplied air mass. The exhaust-gas recirculation rate is regulated in gasoline and diesel engines by means of the air-mass flow.

The maximum air-mass flow to be measured lines in the (time) average within the range of 400 to 1,200 kg/h. As a result of the modest idle requirements of modern engines, the ratio between minimum and maximum air flow is 1:90 to 1:100 in gasoline engines and due to the higher idle requirements 1:20 to 1:40 in diesel engines. Because of stringent exhaust-emission and consumption requirements, levels of accuracy of 1 to 2% from the measured value must be achieved. Referred to the measurement range, this can mean an – unusually high for motor vehicles – measurement accuracy of 10^{-4}.

The air is not drawn in continuously by the engine, but rather in time with the opening of the intake valves. For this reason, the air flow pulses strongly – particularly with the throttle valve wide open – at the measuring point (Figure 1). The measuring point is always located in the intake tract between the air filter and the throttle valve. Due to resonances of the intake manifold, pulsation of the intake manifold – above all in 4-cylinder engines in which the induction or charge phases do not overlap – is so strong that even brief reverse flows occur. These must be correctly detected by an accurate flow meter.

Figure 1: Qualitative characteristic of intake-air pulsation in a 4-cylinder gasoline engine

Operating point: n = 3,000 rpm, full load.
Q_L Mean air flow rate.



Measurement principles and applications

Pitot-tube flow meters

A medium of uniform density ρ at all points flows through a tube with a constant cross-section A at a velocity v which is virtually uniform in the tube cross-section (inlet flow). The following applies:

(1) Volume flow rate $Q_V = v\,A$,
(2) Mass flow rate $Q_M = \rho\,v\,A$.

If an orifice plate is then installed for flow measurement in the flow duct, forming a restriction, this will result in a pressure differential (dynamic pressure) Δp in accordance with Bernoulli's Law. The following applies:

(3) $\Delta p = \text{const} \cdot \rho\,v^2$,

where the constant is dependent on the tube and the orifice-plate cross-sections. This pressure difference can be measured directly with a differential-pressure sensor or as a force acting on a sensor plate. Pivoting, variable-position sensor flaps (Figure 2) leave a variable section of the flow cross-section unobstructed, with the size of the free diameter being dependent on the flow rate. The sensor flap is pressed as the flow increases against a mostly constant counterforce. A potentiometer monitors the characteristic flap positions for the respective flow rates.

Measurement errors can occur in cases where the sensor-plate mechanical inertia prevents it from keeping pace with a rapidly pulsating air current (full-load condition at high engine speeds).

If the density ρ changes due to temperature fluctuations or the altitude, the measurement signal changes with $\sqrt{\rho}$. An air-temperature sensor and a barometric pressure sensor must be used to compensate for this.

The higher sensitivity of the output signal required for a large measurement range with small air masses is obtained from the mechanical and electrical design of the air-flow sensor used in for example the earlier L-Jetronic gasoline injection system.

This method of flow measurement has not been used in new developments for some time now.

Hot-wire air-mass meters

Hot-wire air-mass meters operate without mechanically moving parts. The hot wire is made from for example platinum whose electrical resistance R increases with temperature. The wire heats up when a current I_H is passed through it. It is cooled when it is in an air flow. The cooling process reduces the resistance of the wire and increases the electric current to produce an equilibrium between the input power P_{el} and the power P_V output by the flow:

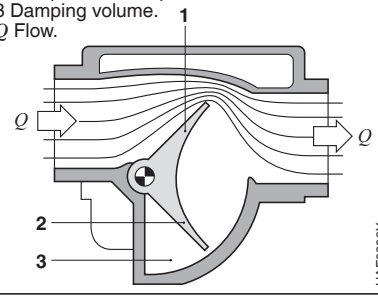
$$(4)\ P_{el} = I_H^2\, R = P_V = c_1\, \lambda\, \Delta T.$$

The dimensions of the hot wire and its air flow assume the constant c_1. The heat conduction λ is approximately proportional to the root of the mass flow ($\sqrt{Q_M}$). Furthermore, if thermal convection is taken into consideration with the medium at rest (no air flow) with the coefficient c_2, the following equation is obtained for the heating current I_H:

$$(5)\ I_H = c_1 \cdot \sqrt{\left(\sqrt{Q_M} + c_2\right)} \cdot \sqrt{\frac{\Delta T}{R}}\ .$$

Figure 2: Pitot-tube flow meter (air-flow sensor)

1 Sensor flap,
2 Compensation flap,
3 Damping volume.
Q Flow.



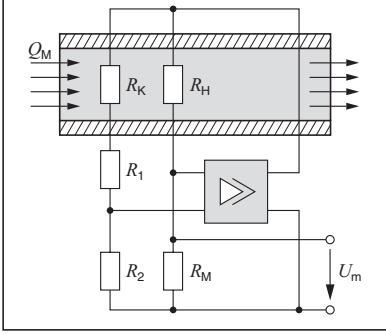
The closed-loop control circuit in the sensor housing (Figure 3, the controller is shown as a box) maintains the heater element (platinum hot wire) at a constant overtemperature in relation to the air temperature. The air temperature is taken into consideration with the compensation resistor R_K, whose resistance value is dependent on the air temperature. The current required for heating provides an extremely precise – albeit nonlinear – index of air-mass flow rate. The associated ECU generally converts the signal into linear form and performs other signal-evaluation tasks, with the air-mass meter signal being sensed in millisecond intervals. Due to its closed-loop design, this type of flow meter can also monitor fast flow-rate variations in the millisecond range since, because of the constant overtemperature of the heater element, its enthalpy (heat content) does not have to be changed by means of time-consuming heat transfers.

However, the flow direction is not recognized with this method, and for this reason these sensors sometimes exhibit substantial deviations when strong pulsations occur in the intake manifold.

To ensure stable and reliable performance throughout an extended service life, the system must burn off all accumulated deposits from the hot wire's surface

Figure 3: Electronic control of hot-wire air-mass meter

Q_M Mass flow,
U_m Measurement voltage,
R_H Hot wire,
R_K Compensation resistor,
R_M Measuring resistor,
R_1,R_2 Trimming resistor.



at approx. 1,000 °C after each phase of active operation (when the ignition is switched off).

The hot-wire air-mass meter was used in earlier engine-management systems (Motronic) and has since been replaced by the hot-film air-mass meter (HFM).

Hot-film air-mass meters using thick-film technology

The first hot-film air-mass meter still manufactured using thick-film technology (HFM2 from Bosch) operates according to the same principle as the hot-wire air-mass meter, but combines all the measuring elements and the control electronics on a single substrate. The flat-design heating resistor is located on the back of the base wafer, with the corresponding temperature sensor on the front. The greater thermal inertia of the ceramic wafer results in somewhat greater response lag than that associated with the hot-wire air-mass meter. The compensation resistor and the heater element are thermally decoupled by means of a laser cut in the ceramic substrate. More favorable air-flow characteristics make it possible to dispense with the hot-wire meter's burn-off decontamination process.

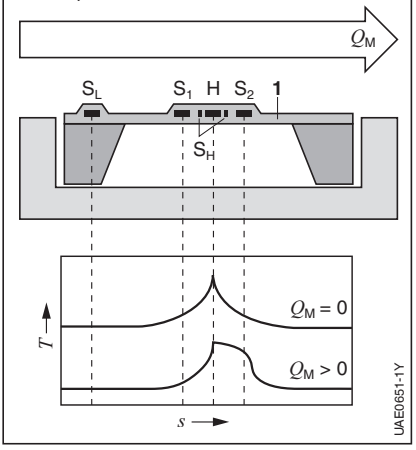
Micromechanical hot-film air-mass meters

Extremely compact micromechanical hot-film air-mass meters (sensor types from Bosch from HFM5 upwards, Figure 4) also operate according to thermal principles. Here the heating and measuring resistors are in the form of thin platinum layers sputtered (vapor-deposited) onto the silicon chip acting as substrate. Thermal decoupling from the mounting is obtained by installing this chip in the area of the heating resistor H on a micromechanically thinned section of the substrate (similar to a pressure-sensor diaphragm). The adjacent heater-temperature sensor S_H and the air-temperature sensor S_L (on the thick edge of the silicon chip) maintain the heating resistor H at a constant overtemperature. This method differs from earlier techniques in dispensing with the heating current as an output signal. Instead, the signal is derived from the temperature difference in the diaphragm detected by the two temperature sensors S_1 and S_2. Temperature sensors are located in the flow path upstream and downstream from the heating resistor H.

Without air inflow (Q_M = 0) the temperature profile is the same on both sides of the heating zone; for the temperature at the measuring points $T_1 = T_2$ applies (Figure 4). If, on the other hand, air flows through the sensor measuring cell ($Q_M > 0$), the area upstream of the heater element is cooled by the cold air and a lower temperature is measured at S_1. Downstream of the heater element the air flowing past is heated by the heater element; a higher temperature is therefore measured at S_2. The temperature difference is, irrespective of the absolute temperature of the air flowing past, a measure of the mass of the air flow. Although (as with the earlier process) the response pattern remains nonlinear, the fact that the initial value also indicates the flow direction represents an improvement over the former method using the heating current.

Figure 4: Sensor element of micromechanical hot-film air-mass meter

1 Dielectric diaphragm.
H Heating resistor,
S_H Heater-temperature sensor,
S_L Air-temperature sensor,
S_1,S_2 Temperature sensors (upstream, downstream)
Q_M Air-mass flow,
s Measuring location,
T Temperature.



In the HFM5 the evaluation electronics integrated in the sensor convert the temperature difference measured at S_1 and S_2 with the aid of an analog circuit into an analog voltage signal between 0 V and 5 V (Figure 6). In the HFM6 and HFM7 the signal is processed with the aid of digital electronics to achieve greater accuracy and a frequency signal is generated. The HFM7 is a further-developed version of the HFM6 with improved electronics and the possibility of outputting an analog signal as an alternative to the frequency signal. This voltage or the frequency is converted in the engine ECU via the sensor curve stored in the program into an air-mass flow. The HFM8 can also output the conditioned measurement signal via the SENT digital bus interface. The accuracy of measured-value acquisition could also be increased with the HFM8.

The hot-film air-mass meter projects with its housing into a measuring tube (Figure 5). A flow rectifier mounted on the measuring tube (e.g., a wire grill) ensures a uniform flow in the measuring tube. The measuring tube is installed after the air filter in the intake tract.

Because of the small size of the measuring element, this flow meter is a partial flow meter, which detects only a specific and very small part of the total flow. The constancy and reproducibility of this division factor have a direct impact on the accuracy of the sensor. A calibration establishes the interrelationship between the air mass Q_M flowing through the measuring tube and the measurement signal generated from the partial air flow.

The inlet and outlet to the micromechanical measuring element are designed and optimized in such a way that heavier particles, such as dust particles and liquid droplets, do not approach the measuring element directly, but are rather diverted ahead of it. Further modifications to the measuring passage resulted in improved protection against contamination directly upstream of the sensor element.

Figure 5: Section through hot-film air-mass meter

1 Electrical connections,
2 Measuring-tube or air-filter housing wall,
3 Evaluation electronics (hybrid circuit),
4 Sensor measuring cell, 5 Sensor case,
6 Partial-flow measuring passage,
7 Outlet, partial air flow Q_M,
8 Inlet, partial air flow Q_M.

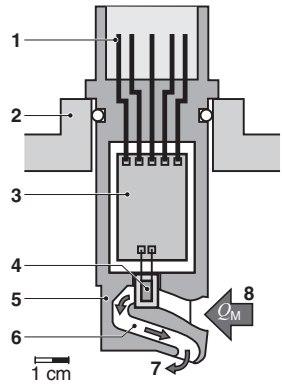
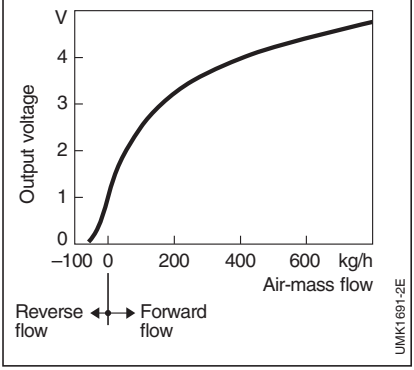


Figure 6: Characteristic curve of hot-film air-mass meter



Acceleration and vibration sensors

Measured variables
The acceleration values a to be measured in motor vehicles are frequently expressed as a multiple of gravitational acceleration g ($1\ g \approx 9.81\ \mathrm{m/s^2}$). Acceleration and vibration sensors are used in the following applications:
– Triggering restraint systems such as e.g. airbag and seat-belt pretensioner (35…100 g).
– Side-impact and upfront sensing (100…500 g).
– Rollover detection (3…7 g).
– Sensing acceleration by the vehicle for the antilock braking system (ABS) and driving-dynamics control (0.8…1.8 g).
– Evaluating bodywork acceleration for chassis control systems: body acceleration (1…2 g), axle and damper (10…20 g).
– Detecting changes in vehicle inclination for anti-theft alarm systems (approx. 1 g).
– Knock control in gasoline engines (measurement range up to 40 g).

Measurement principles
Acceleration sensors measure the force exerted by an acceleration a on an inert mass m:

(1) $F = m\,a$.

There are both position- and (mechanically) stress-measuring systems.

Position-measuring systems
Position-measuring systems are used particularly in the range of very small accelerations. All acceleration sensors are spring-bound right down to the gravitational pendulum. In other words, the inert mass is elastically connected to the body whose acceleration a is to be measured (Figure 1a).

Excursion-measuring sensor
In the static case the acceleration force is in equilibrium with the restoring force applied to the spring which has been deflected by x:

(2) $F = m\,a = c\,x$
where c = spring constant.

The resulting excursion is converted by means of a suitable measurement method (e.g. piezoelectric, capacitive, piezoresistive or thermal) into an electrical signal.

The system's measurement sensitivity S is therefore:

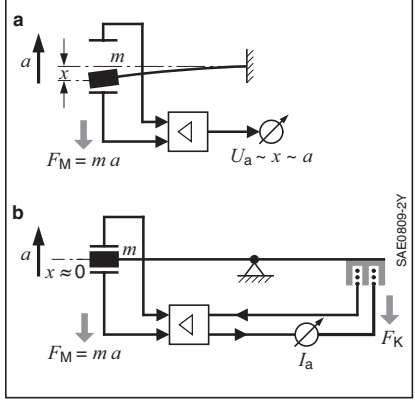
(3) $S = \frac{x}{a} = \frac{m}{c}$.

In other words, a large mass together with low spring stiffness (or constant) result in high measurement sensitivity.

In the dynamic case a damping force and an inertial force must be taken into account as well as the spring force. The damping force is proportional to the velocity $\dot{x}$ and is described with the damping coefficient p. The inertial force is proportional to the acceleration $\ddot{x}$. The oscillatory (resonant) system is described by the following equation:

(4) $F = m\,a = c\,x + p\,\dot{x} + m\,\ddot{x}$.

Figure 1: Displacement-measuring acceleration sensors
a) Excursion-measuring acceleration sensor,
b) Position-controlled acceleration sensor.
a Measurement acceleration,
x System excursion,
F_M Measuring force (inertial force on mass m),
F_K Compensation force,
I_A Output current,
U_A Output voltage.



If negligible friction ($p \approx 0$) is presumed, the system resonates at the natural frequency ω_0:

(5) $\omega_0 = \sqrt{\frac{c}{m}}$.

Thus, according to equation (3), the measurement sensitivity S is directly linked to the resonant frequency ω_0 in the following manner:

(6) $S\,\omega_0^2 = 1$.

In other words, it can be expected that sensitivity is reduced to a factor of ¼ when the resonant frequency is increased by a factor of 2. It is only below their resonant frequency that such spring-mass systems display adequate proportionality between measured variable and excursion.

Position-controlled sensor

Position-measuring systems also permit the use of the compensation method in which the system excursion caused by acceleration is compensated for by an equivalent restoring force (Figure 1b). In practical terms the system then ideally always operates very close to its zero point (high linearity, minimum cross sensitivity, high temperature stability). These position-controlled systems also have, on account of their control, higher stiffness and cutoff frequency than excursion systems of the same kind. Any lack of mechanical damping can be electronically generated here.

System behavior

In non-position-controlled systems the damping and resonance of the systems play key roles in system behavior. The application of Lehr's damping ratio D has proved effective in describing this behavior:

(7) $D = \frac{p}{m} \cdot \frac{1}{2\omega_0}$.

Equation (4) can then be formulated as

(8) $\frac{F}{m} = \ddot{x} + 2\,\omega_0 D\,\dot{x} + \omega_0^2 x$.

Using the dimensionless variable D permits a simple comparison to be made of different oscillatory systems: transient response and resonant response are defined to a large extent by this damping ratio.

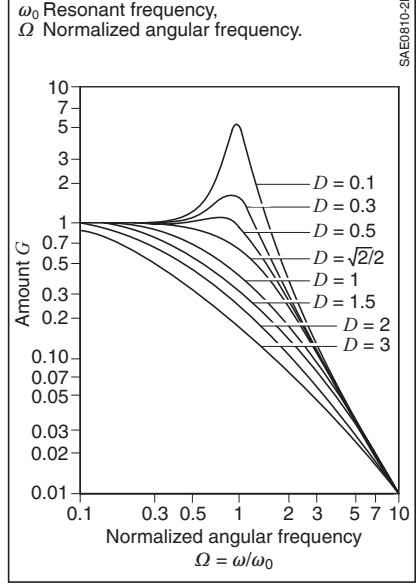
With very low damping ($D \rightarrow 0$) the system exhibits a resonance sharpness at ω_0. This is often unwelcome in that on the one hand it can lead to the system being destroyed and on the other hand it results in a nonuniform transfer function ($G >> 1$).

As damping factors increase, the system's resonance sharpness continues to decrease. While no more resonance sharpness is obtained under periodic excitation for damping factors $D = \sqrt{2}/2$ (Figure 2), any oscillating transient condition in the case of a step excitation disappears for factors $D > 1$.

In order to achieve as uniform a system response as possible in the region $G \approx 1$ over a wide frequency range, it is important when designing the system to define its damping precisely. Ideally this is influenced for example only slightly by

Figure 2: Amplitude-resonance curve

G Transfer function,
D Damping,
ω Angular frequency,
ω_0 Resonant frequency,
Ω Normalized angular frequency.



fluctuations in temperature. In practice sensors are therefore often designed in the range D = 0.5…1.0.

Mechanical stress-measuring systems
The piezoelectric effect can also be utilized to measure an acceleration. In this case the exertion of force by the external acceleration causes a mechanical stress in the piezoelectric material. These piezoelectric materials generate a charge Q under the influence of a force F on its surfaces, which are fitted with electrodes (Figure 3). This charge is proportional to the mechanical stress generated by the force.

For use as a sensor the generated charges are discharged via the external resistor of a measuring circuit or via an internal resistor of the sensor. The sensor therefore can measure dynamically, but not statically. The typical cutoff frequency is above 1 Hz.

The piezoelectric effect is utilized for example in piezoceramic bimorphous spring elements and in piezoelectric knock sensors.

Thermal acceleration sensors
These sensors generate a confined heated gas zone (Figure 4). The gas in this zone has a lower density than the surrounding cooler gas. During an acceleration the low-density gas zone moves within the surrounding cooler gas. The resulting asymmetry can be sensed by temperature sensors which are interconnected in a bridge circuit. The bridge voltage constitutes the acceleration signal.

This principle is used in numerous applications such as for example rollover sensing, the driving-dynamics control, and even in smart phones.

Figure 3: Piezoelectric effect
a) Longitudinal effect,
b) Transversal effect,
c) Shear effect.
F Force,
Q Charge.

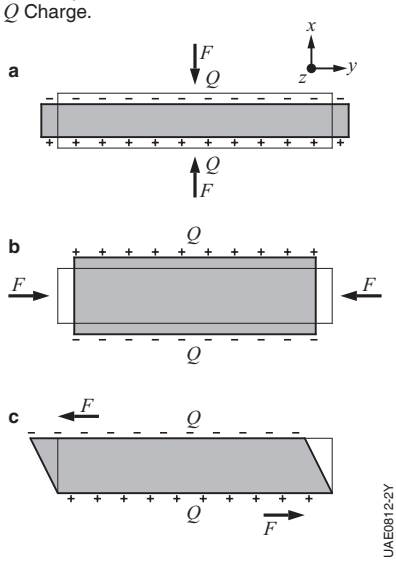
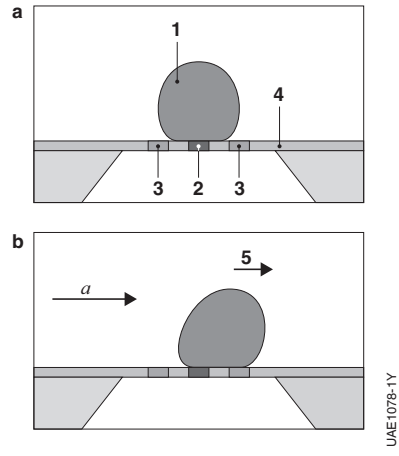


Figure 4: Principle of thermal acceleration sensors
a) Acceleration a = 0,
b) Acceleration a > 0.
1 Heated gas zone,
2 Heater element,
3 Temperature sensors,
4 Substrate layer,
5 Displaced hot zone.



Applications

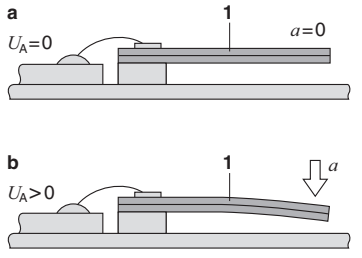
Piezoelectric acceleration sensors

Piezoelectric bimorphous spring elements or two-layer piezoceramics (Figure 5) are used for restraint-system sensors for triggering seat-belt tighteners, airbags, and rollover bars. Their intrinsic mass causes them to deflect under acceleration to provide a dynamic (not DC response pattern) signal with excellent processing characteristics (typical cutoff frequency: 10 Hz). The sensor element is located in a sealed housing shared with the initial signal-amplification stage. It is sometimes encased in gel for physical protection. It consists of two oppositely polarized piezoelectric layers which are bonded to each other (bimorphous). An acceleration acting on it causes a mechanical tensile stress ($\varepsilon > 0$) in one layer and a compressive stress ($\varepsilon < 0$) in the other layer. The metallic layers on the upper and lower sides of the spring element serve as electrodes via which the resulting voltage is tapped.

This structure is packed together with the evaluation electronics in a hermetically sealed housing. The electronic circuit consists of an impedance transformer and an adjustable amplifier with specified filter characteristic.

The sensor's actuating principle can also be inverted: The sensor can be checked with an additional actuator electrode (on-board diagnosis).

Figure 5: Piezoelectric sensor
a) At rest,
b) Subject to acceleration a.
1 Piezoceramic bimorphous spring element.
U_A Measurement voltage.

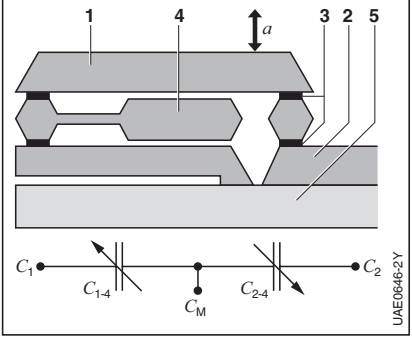


Micromechanical bulk silicon acceleration sensors

The first generation of micromechanical sensors relied on anisotropic and selective etching techniques to fabricate the required spring-mass system from the full silicon wafer (bulk silicon micromechanics) and produce the spring profile (Figure 6).

Capacitive taps have proven especially effective for the high-precision measurement of this seismic-mass deflection. This design entails the use of supplementary silicon or glass wafers with counter-electrodes above and below the spring-held seismic mass. This leads to a 3-layer structure, whereby the wafers and their counter-electrodes also provide overload protection. This layout corresponds to a series connection of two capacitors ($C_{1\text{-}M}$ and $C_{2\text{-}M}$). Alternating voltages are supplied at the connections C_1 and C_2, the superposition of these voltages is tapped at C_M, i.e. at the seismic mass. In the rest state the capacitances $C_{1\text{-}M}$ and $C_{2\text{-}M}$ are ideally equal. When an acceleration a acts in the measurement direction, the silicon center wafer is deflected as the seismic

Figure 6: Bulk silicon acceleration sensor
1 Silicon upper wafer,
2 Silicon lower wafer,
3 Silicon oxide,
4 Silicon center wafer M (seismic mass),
5 Glass substrate.
a Acceleration,
C Measurement capacitances.



mass. The change in distance between the upper and lower wafers gives rise to a change in capacitance in the capacitors $C_{1\text{-M}}$ and $C_{2\text{-M}}$, and thus a difference ΔC that is proportional to the applied acceleration. This in turn changes the electrical signal at C_M, which is amplified and filtered in the evaluation electronics.

Filling the sensor's hermetically sealed oscillatory system with a precisely metered charge of air leads to a very space-saving, inexpensive form of damping which also exhibits low temperature sensitivity.

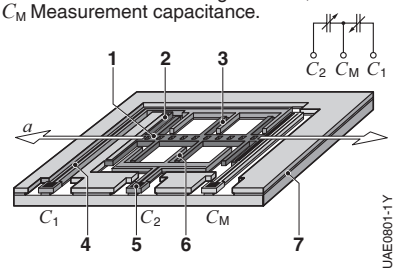
This type of sensor is mainly used for low-level accelerations (< $2g$) and relies upon a two-chip concept (sensor chip + CMOS processing chip).

Surface-micromechanical acceleration sensors

Surface-micromechanical sensors (SMM) have more compact dimensions than bulk silicon sensors. In SMM technology an additive process is employed to construct the spring-mass system on the surface of the silicon wafer (Figure 7). In the sensor core the seismic mass with its comb-shaped electrodes is connected by spring elements to anchor points of the silicon oxide. The chip features fixed, also comb-shaped electrodes on both sides of these moving electrodes. The fixed and moving electrode fingers create individual capacitances which are connected in parallel. This produces an effective total capacitance of 300 fF up to 1 pF. Two rows of electrode fingers connected in parallel produce two useful capacitances (C_1-C_M and C_2-C_M), which change in opposite directions when the seismic mass is deflected. The spring-mass system is subjected during an acceleration to a deflection which behaves linearly to the acceleration via the spring restoring force. An electrical output signal that is linearly dependent on the acceleration can be obtained by evaluation of this differential capacitor.

Figure 7: Surface-micromechanical acceleration sensor with capacitive tap

1 Spring-mounted seismic mass with electrodes,
2 Spring,
3 Fixed electrodes with capacitance C_1,
4 Printed aluminum conductor (for self-test function),
5 Bond pad,
6 Fixed electrodes with capacitance C_2,
7 Silicon oxide.
a Acceleration in sensing direction,
C_M Measurement capacitance.



Because of the low capacitance of typically 1 pF the evaluation electronics are therefore integrated together with the sensor on the same chip, or very closely connected to it on the same substrate or leadframe. The raw signal is amplified, filtered, and conditioned for the output interface. Analog voltages, pulse-width-modulated signals, SPI protocols (Serial Peripheral Interface) or PSI5 protocols (current interfaces, see PSI5) are usual for this purpose.

The mechanical and electrical signal paths of the sensor can be checked with a self-test function. The sensor structure is deflected by means of an electrostatic force (simulation of an acceleration) and the resulting sensor signal is compared with a setpoint value.

Depending on the applications, the sensor elements are designed for different measurement ranges between 1 g and 500 g. Initially these sensors were used for high acceleration rates (50…100 g, for passenger-protection systems), but are now also used for lower acceleration rates (e.g. for the driving-dynamics control).

Piezoelectric knock sensor

The knock sensor is in principle a vibration sensor. It senses structure-borne sound vibrations that occur as "knocking" in a gasoline engine during uncontrolled combustions (Figure 7). The structure-borne sound vibrations must be able to be introduced from the measuring location on the engine block undamped and without resonance into the knock sensor. A suitable measuring point and a fixed screw connection are required for this purpose.

The knock sensor is screwed to the engine block (Figure 8). A seismic mass exerts as a result of its inertia pressure forces in time with the exciting vibrations on an annular piezoceramic element. These forces induce a charge transfer within the ceramic element. This generates between the upper and lower sides of the ceramic element a voltage that is tapped via contact disks. The knock signal generated in the sensor is forwarded to the engine ECU, where it is conditioned and evaluated.

When knocking noises are detected the engine ECU retards the ignition timing and thus counteracts any further knocking (see Knock control). The vibration frequency is typically 5 to 25 kHz.

Figure 8: Knock sensor (design and mounting)

1 Annular piezoceramic element,
2 Seismic mass with pressure forces F,
3 Housing,
4 Screw,
5 Contact,
6 Electrical connection,
7 Engine block.
V Vibration.

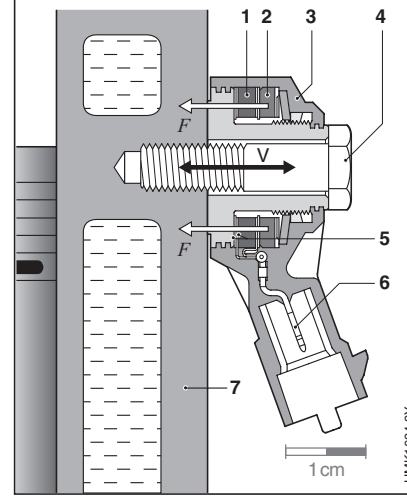
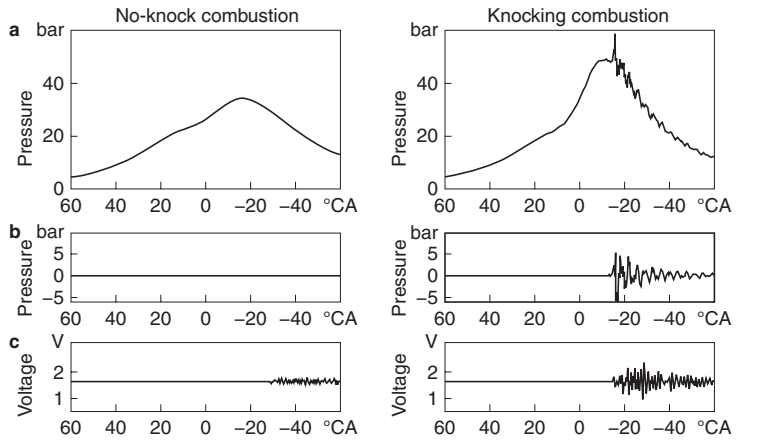


Figure 7: Pressure characteristic in the combustion chamber and corresponding knock-sensor signals

a) Typical characteristic of combustion-chamber pressure (measured on a test engine),
b) Bandpass-filtered signal of combustion-chamber pressure,
c) Structure-borne noise signal sensed by the knock sensor.



Pressure sensors

Measured variables
The measured variable pressure is defined as a nondirectional force acting in all directions which occurs in gases and liquids. It still propagates very well in gelatinous substances and soft sealing compounds. The most important measured pressures in a motor vehicle are:
– Intake-manifold pressure and charge-air pressure (1 to 6 bar),
– Ambient pressure (approx. 1 bar), e.g. for boost-pressure control,
– Vacuum in the brake booster (approx. 1 bar relative to atmosphere),
– Differential pressure at the diesel particulate filter for detecting the filter-load state and leaks (up to 1 bar differential pressure),
– Brake pressure (10 bar) in electropneumatic brakes,
– Air-spring pressure (16 bar) for vehicles with air suspension,
– Tire pressure (5 bar absolute) for monitoring tire pressure,
– Hydraulic reservoir pressure (approx. 200 bar) for the antilock braking system and power steering,
– Shock-absorber pressure (200 bar) for chassis-control systems,
– Refrigerant pressure (35 bar) in air-conditioning systems,
– Modulation pressure (35 bar) in automatic transmissions,
– Fluid pressure for clutch actuation in double-clutch transmissions (20 bar),
– Engine-oil pressure for demand-controlled oil pumps (10 bar),
– Brake pressure in brake master and wheel-brake cylinders (200 bar) and for automatic yaw-moment compensation in electronically controlled brakes,
– Positive and vacuum pressure in fuel tank (0.5 bar) for on-board diagnosis,
– Combustion-chamber pressure (100 bar, dynamic) for misfiring and knock detection,
– Pump-side pressure in diesel distributor-type injection pumps (up to 1,000 bar, dynamic) for electronic governing,
– Fuel pressure in the low-pressure circuit for demand-controlled fuel pumps (diesel and gasoline; 10 bar),
– Rail pressure in LPG and CNG systems (4 to 16 bar),
– Rail pressure for diesel common rail (up to 2,700 bar),
– Rail pressure for gasoline direct injection (up to 280 bar),
– Oil pressure for taking into account the engine load in the service display.

Measurement principles
Dynamically and statically acting pickups or sensors are used to measure pressures. Dynamically acting pressure sensors include for example microphones, which are insensitive to static pressures and are only used to measure pressure pulsations in gaseous or liquid media. Since up to now, practically only static pressure sensors have been used in automobiles, these will be dealt with in more detail here. Static pressure measurement is effected directly by the deformation of a diaphragm.

Diaphragm-type sensors
Design
The most common method used for pressure measurement (also in automotive applications) uses a thin diaphragm as the intermediate stage. The pressure to be measured is first of all applied to one side of this diaphragm so that this bends to a greater or lesser degree as a function of the pressure (Figure 1a). Within a very wide range, its diameter and thickness can be adapted to the particular pressure range. Low-pressure measuring ranges lead to comparatively large diaphragms which can deform in the range from 0.01 to 1 mm. Higher pressures demand thicker, low-diameter diaphragms which only deform very slightly by a few micrometers.

To be precise, a diaphragm's deformation depends upon the difference in the pressure applied to its top and bottom sides. Accordingly, there are three different basic types of pressure sensor:
– Absolute-pressure sensors,
– Differential-pressure sensors,
– Relative-pressure sensors.

Relative-pressure sensors measure the differential pressure relative to the ambient pressure.

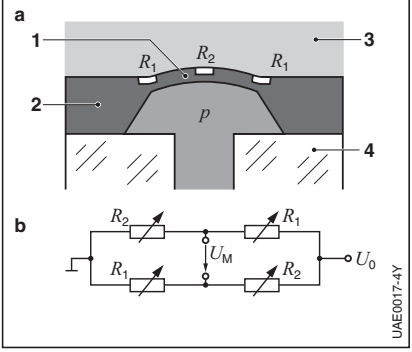
Voltage-measuring methods dominate in all pressure ranges and here in practical terms only strain-gage technology is used.

Strain-gage technology

The strains on the diaphragm when a diaphragm-type sensor is subjected to deformation are sensed by strain gages (strain-gage resistors) which are applied to the diaphragm (e.g. diffused or vapor-deposited) (Figure 1a). Under the influence of mechanical stresses their electrical resistance changes on account of the piezoresistive effect. The K factor (gage factor) specifies the relative change of the strain-gage resistor referred to the relative change of its length. For the resistors diffused into the monocrystalline silicon in micromechanical pressure sensors this factor is especially high, typically $K \approx 100$.

Figure 1: Semiconductor absolute-pressure sensor

a) Cutaway view,
b) Wheatstone-bridge circuit.
1 Diaphragm,
2 Silicon chip,
3 Reference vacuum,
4 Glass (Pyrex),
p Measured pressure,
U_0 Supply voltage,
U_M Measurement voltage,
R_1 Strain-gage resistor (compressed),
R_2 Strain-gage resistor (stretched).



The resistors are interconnected to form a Wheatstone bridge (Figure 1b). The diaphragm is deformed to varying degrees depending on the effective pressure; in the process two resistors are stretched and two resistors are compressed. This alters their electrical resistance and thus also the measurement voltage (bridge voltage). The measured voltage is a measure of the pressure. This circuit produces a higher measurement voltage than is the case when a single resistor is evaluated. The Wheatstone bridge thus provides for a highly sensitive sensor.

Applications

Micromechanical pressure sensors

Micromechanical pressure sensors are used for pressure ranges of less than 6 bar (in low-pressure sensors) and for pressure ranges of less than 70 bar (in medium-pressure sensors).

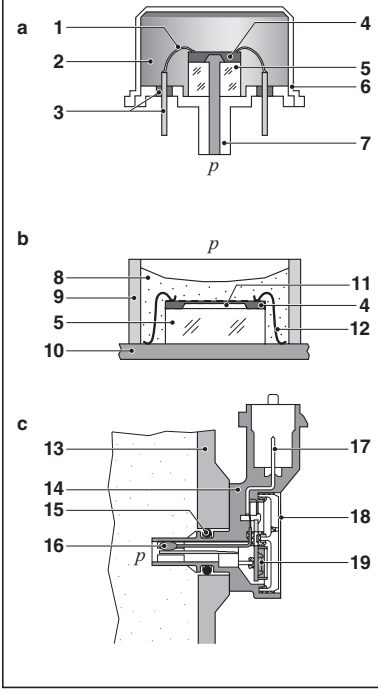
The measuring cell of the micromechanical pressure sensor consists of a silicon chip in which a thin diaphragm is micromechanically etched (Figure 1). Four resistors arranged in a Wheatstone-bridge circuit are diffused on the diaphragm (see above). The bridge voltage is a measure of the pressure on the diaphragm.

The bridge signal must still be linearized and amplified. Temperature effects must also be compensated. This is performed either in a circuit integrated on the measuring chip or in a separate ASIC (application-specific integrated circuit). The output signal can be transferred either in the form of an analog voltage (0 to 5 V) or in digital form, e.g. via the SENT interface.

The advantage of digital transfer protocols over analog signals is that interface tolerances (contact and cable resistances) play no role and that as a rule additional information (fault codes, temperature signals, etc.) can still be transferred.

In the case of sensors for very aggressive or liquid measured media (e.g. for fuel-pressure and charge-air pressure sensor) “reverse assembly” is often used in which the measured pressure is conducted to an electronically passive cavity recessed into the side of the sensor

Figure 2: Absolute-pressure sensor
a) Reverse assembly
(reference vacuum on structure side),
b) Simplified assembly with protective gel,
c) "Outer packaging"
for intake-manifold installation
with integrated air-temperature sensor.
1,3 Electrical connections
with glass-enclosed lead-in,
2 Reference-vacuum chamber,
4 Measuring cell (chip)
with evaluation electronics,
5 Glass base,
6 Cap,
7 Supply for measured pressure p,
8 Protective gel,
9 Gel frame,
10 Ceramic hybrid,
11 Cavity with reference-vacuum chamber,
12 Bonded connection,
13 Manifold wall,
14 Housing,
15 Sealing ring,
16 Temperature sensor,
17 Electrical connection,
18 Housing cover,
19 Measuring cell.
p Measured pressure.



chip (Figure 2a). For maximum protection, the – much more sensitive – side of the chip with the measuring resistors, evaluation electronics and contacts is enclosed in a reference-vacuum chamber located between the housing's base and the soldered metal cap.

However, a cheaper arrangement is to seat the silicon chip with etched diaphragm and four strain-gage resistors as a measuring cell on a glass base. The reference-vacuum chamber is located in a cavity between the diaphragm and the glass base (Figure 2b). Alternatively, measuring chips made of porous silicon with an enclosed reference-vacuum chamber can be used for absolute-pressure sensors. The silicon chip is subjected to pressure from the side on which the measuring resistors and the evaluation electronics are located. This sensitive side of the chip is protected by a suitable gel against the pressure medium. The compact design of integrated single-chip sensors means that they are suitable for mounting directly on the intake manifold (Figure 2c).

Low-pressure sensors
This design is used in virtually all low-pressure sensors (maximum pressure less than 6 bar) such as for example charge-air pressure sensors and barometric sensors. For differential-pressure sensors (e.g. for diesel particulate filters or brake boosters) both sides of the diaphragm are subjected to the pressure medium.

These sensors can also be used in tire-pressure monitoring systems. Measurement will be continuous and non-contacting.

Medium-pressure sensors
The range of medium-pressure sensors is essentially described by a maximum pressure-measuring range of approx. 6 to 70 bar. These sensors are used to measure engine-oil pressure, fuel pressure ahead of the high-pressure pump, hydraulic pressure in automatic transmissions (torque-converter, double-clutch, and CVT transmissions), to measure pressure in LPG and CNG systems, and to measure pressure in air-conditioning systems.

The wide variety of applications means that there is a broad range of mechanical and electrical interfaces. The sensors are screwed into lines or into the crankcase and transmission case. A metallic taper seal has proven effective, but O-rings and sealing rings are also used. The design of the medium-pressure sensor very much follows that of the high-pressure sensor, with the metal diaphragm being replaced by a silicon-micromechanics-technology sensor element. In this case the sensor elements used in low-pressure sensors have been adapted by optimized etching methods for higher pressures and increased bursting-pressure strength. Here too the measurement signal of a Wheatstone-bridge circuit is evaluated. The sensor signal is conditioned and issued as an analog or a digital output signal. This is done either integrated in a silicon chip (1-chip technology) or by measuring element and evaluation ASIC (2-chip technology).

Combined sensors for pressure and temperature are used in some applications (e.g. in the engine oil-level sensor). For this purpose an NTC is integrated in the threaded fitting and the thermistor of the NTC is evaluated in the ECU by suitable circuits. The SENT protocol offers the opportunity to issue the pressure and temperature signals in a joint output signal. This dispenses with the need for a separate line for the temperature signal.

High-pressure sensors with metal diaphragm

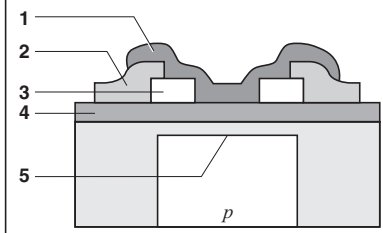
Diaphragms made of high-quality spring steel are used at very high pressures as have to be measured for example in the rail of the common-rail system or gasoline direct injection for closed-loop control (Figures 3 and 4). Further applications include braking systems such as the anti-lock braking system and driving-dynamics control, and also industrial hydraulics.

Four thin-film strain gages are applied to the diaphragm; the measurement principle is the same as for micromechanical pressure sensors. The steel diaphragm insulates the measured medium, differs from silicon in retaining a yield range for enhanced burst resistance, and is easy to install in metallic housings.

Insulated sputtered (vapor-deposited) metallic thin-film strain gages ($K \approx 2$) and even polysilicon strain gages ($K \approx 40$) offer permanently high sensor accuracy. Amplification, calibration, and compensation elements can be combined in a single ASIC, which is then integrated together with the required EMC protection on a small carrier in the sensor housing.

Figure 3: High-pressure sensor with metal diaphragm, measuring element

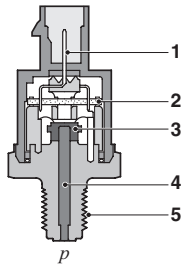
1 SiN_x passivation,
2 Gold contact,
3 Polysilicon strain-gage resistor,
4 SiO_2 insulation,
5 Steel diaphragm.
p Measured pressure.



UAE0650-1Y

Figure 4: High-pressure sensor with metal diaphragm, design

1 Electrical connection (plug),
2 Evaluation circuit,
3 Steel diaphragm with strain-gage resistors,
4 Pressure connection,
5 Fastening thread.
p Measured pressure.



UMK1576-3Y

Temperature sensors

Measured variables

Temperature is defined as a nondirectional quantity which characterizes the energy state of a given medium, and which can be a function of time and location.

(1) $T = T(x, y, z, t)$,
x, y, z Space coordinates,
t Time,
T measured according to the Celsius or Kelvin scale.

In the case of gaseous and liquid measured media, it is generally possible to measure temperature at all points without any problems; in the case of solid bodies, measurement is usually confined to the surface. The temperature sensors mostly commonly used need to have direct, close contact with the measured medium (contact thermometer) to enable them to assume the medium temperature as precisely as possible. Outside temperature, intake-air temperature, and exhaust-gas temperature are examples of temperature measurement of gaseous media. Coolant temperature and engine-oil temperature are examples of temperature measurement of liquid media. Further areas of application of temperature measurement in vehicles are shown in Table 1. They can differ greatly in terms of required measuring accuracy and permissible measuring time.

For special cases non-contact temperature sensors are also used which determine the temperature of a body or a medium on the basis of the (infrared) heat radiation emitted by it (radiation thermometer, pyrometer, thermal imaging camera). These temperature sensors are used in night-vision systems (see Far-infrared systems) and for pedestrian recognition with infrared cameras.

At many locations temperature is also measured as a secondary variable in order that it can be compensated in cases where temperature variations trigger faults or act as an undesirable influencing variable – for example in sensors for other physical variables whose measured value is to depend as little as possible on the temperature.

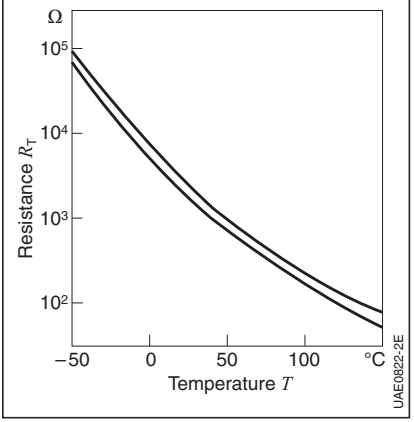
Table 1: Temperature ranges in motor vehicles

Measuring point	Range in °C
Intake and charge air	–40…170
Outside atmosphere	–40…60
Passenger compartment	–20…85
Ventilation & heating air	–20…60
Evaporator (air conditioner)	–10…50
Coolant	–40…130
Engine oil	–40…170
Battery	–40…100
Fuel	–40…120
Tire air	–40…120
Exhaust emissions	100…1,000
Brake calipers	–40…2,000

Measurement principle

Temperature measurements in motor vehicles are conducted almost entirely by exploiting the sensitivity to temperature variation found in electrical resistance materials with a positive temperature coefficient (PTC, NTC thermistor) or with a negative temperature coefficient (NTC, PTC thermistor) as contact thermometers. Depending on the temperature, the sensor exhibits a certain resistance

Figure 1: Temperature curve with limits of variation of an NTC thermistor



value (Figure 1). Conversion of the resistance variation into an analog voltage is performed predominantly by adding a temperature-neutral or inversely sensitive resistance to a voltage divider (Figure 2a) or by supplying an injected current (Figure 2b). For the voltage-divider circuit the following is obtained

$$U_A(T) = U_0 \frac{R(T)}{R(T)+R_V}.$$

With the injected current I_0 the measurement voltage is obtained at

$$U_A(T) = I_0\, R(T).$$

The voltage measured at the measuring resistor is thus temperature-dependent. This analog voltage is digitized in the ECU by an analog-digital converter and then assigned via a characteristic curve to a temperature value.

Alternatively, signal processing can also be performed by an evaluation circuit integrated in the sensor; the temperature value is then transferred via a digital interface (see PSI5).

Sensor types

Sintered-ceramic resistors (NTC)

Semiconducting sintered-ceramic resistors made of heavy-metal oxides and oxidized mixed crystals (sintered in pearl or plate form) display an exponentially falling temperature curve (Figure 1). It can be described to good approximation by the following exponential equation:

$$R(T) = R_0\, e^{B\left(\frac{1}{T}-\frac{1}{T_0}\right)},$$

where
$R_0 = R(T_0)$,
B = 2,000…5,000 K,
T Temperature in K.

The resistance value can vary up to five powers of ten, for example typically from some 100 kiloohms to some 10 ohms. High thermal sensitivity means that applications are restricted to a "window" of approximately 200 K; however, this range can be defined within a latitude of −40 °C to approximately 850 °C by selecting an appropriate NTC. Close tolerances of ±0.5 K at a selectable reference point are possible for example by means of read-out, but this is also reflected in the cost.

Thin-film metallic resistors (PTC)

Thin-film metallic resistors, integrated on a single substrate wafer (e.g. by vapor-depositing or sputtering) with two supplementary, temperature-neutral trimming resistors (Figure 3a), are characterized by extreme precision, as they can be manufactured and then "trimmed" with lasers to maintain exact response-curve tolerances over long periods of time (Figure 3b). Here the values of the trimming resistor connected in series to the measuring resistor and of the trimming resistor connected in parallel are changed.

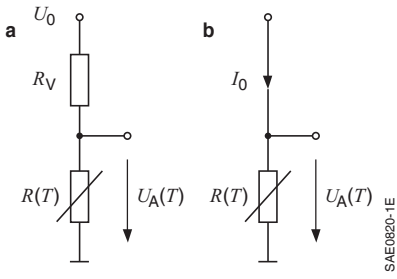
The use of layer technology makes it possible to adapt the substrate (ceramic, glass, plastic foil) and the upper layers (plastic molding or paint, sealed foil, glass and ceramic coatings) to the respective application, and thus provide protection against the monitored medium.

Although metallic layers are less sensitive to thermal variations than the ceramic-oxide semiconductor sensors, both linearity and reproducibility are better. The following formulation is used to provide the computational description:

$$R(T) = R_0\,(1+\alpha\,\Delta T + \beta\,\Delta T^2+\ldots)$$

Figure 2: Resistance-measurement methods

a) Voltage-divider circuit,
b) Measurement with injected current.
U_0 Supply voltage,
I_0 Injected current,
R_V Voltage-divider resistor,
$R(T)$ Measuring resistor,
$U_A(T)$ Measurement voltage.



where
$\Delta T = T - T_0$,
T_0 = 20 °C (reference temperature),
α Linear temperature coefficient,
β Quadratic temperature coefficient.

The coefficient β is small for metals, but not entirely negligible. Therefore in sensors of this type the measurement sensitivity is characterized with a mean temperature coefficient, "TC 100". It corresponds to the mean characteristic-curve gradient between 0 °C and 100 °C (Figure 4) and is obtained from the following relationship:

$$TC\,100 = \frac{R(100\,°C) - R(0\,°C)}{R(0\,°C)\cdot 100\ K} = \alpha_{100}\,.$$

Figure 3: Metal-film temperature resistor
a) Design,
b) Curve trimming.
1 Auxiliary contacts, 2 Bridge.
R_{Ni} Nickel-film resistor,
$R(T)$ Resistor referred to temperature T,
R_P, R_S Trimming resistors (parallel and serial).

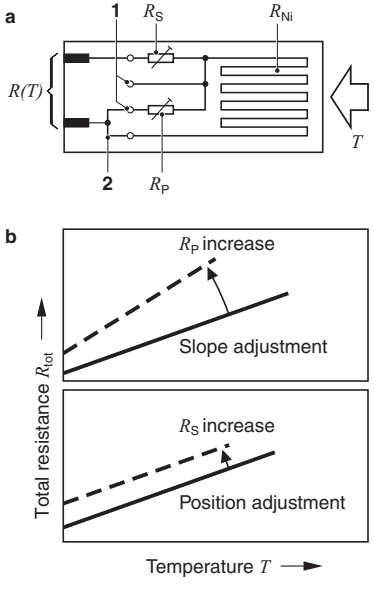


Table 2 shows the mean temperature coefficients for some metals.

Metal-film resistors are generally manufactured and marketed with a basic value (resistance value at 20 °C) of 100 Ω or 1,000 Ω (e.g. under the designation Pt 100 or Pt 1000) in different tolerance classes. Platinum resistors (Pt) do have the lowest temperature coefficient, but are very accurate and resistant to aging.

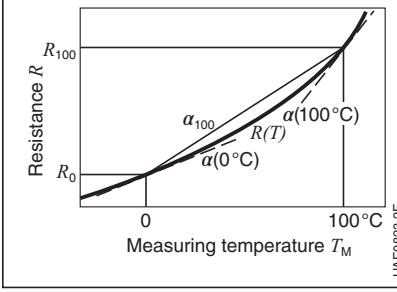
In addition to compensating elements, further active and passive circuit elements for the thin-film sensor can basically be integrated on the carrier chip. This facilitates initial, advantageous signal conditioning and conversion at the measuring point.

Thick-film resistors (PTC and NTC)
Thick-film resistors consist of a carrier chip (e.g. thin ceramic wafer) on which the resistor material is applied as paste (e.g.

Table 2: Mean temperature coefficient of some metals

Sensor material	TC 100	Measurement range
Nickel (Ni)	$6.18\cdot10^{-3}$/K	–60...250 °C
Copper (Cu), value dependent on composition	$3.8\cdot10^{-3}$/K up to $4.3\cdot10^{-3}$/K	–50...200 °C
Platinum (Pt) (DIN EN 60751, [1])	$3.85\cdot10^{-3}$/K	–200...850 °C

Figure 4: Definition of mean temperature coefficient TC 100 = α_{100}



metal oxide) using thick-film technology and sintered at 800 °C to the substrate.

Thick-film resistors with their higher resistivity (low surface-area requirement) and positive and negative temperature coefficients are generally employed as temperature sensors for compensation purposes. They have nonlinear response characteristics (without, however, the extreme variations of the massive NTC resistor) and can for example be laser-trimmed. The measurement effect can be enhanced by using NTC and PTC materials to form voltage-divider circuits.

Monocrystalline silicon semiconductor resistors (PTC)

When monocrystalline semiconductor materials, such as silicon, are used to manufacture the temperature sensor, it is possible to integrate additional active and passive circuitry on the sensor chip. This facilitates initial signal conditioning at the measuring point. Due to the closer tolerancing which is possible, the measuring element is manufactured according to the "spreading-resistance" principle. The current flows through the measuring resistor and through a surface-point contact before arriving at the silicon bulk material. It then proceeds, widely distributed, to a counter-electrode covering the base of the sensor chip (Figure 5).

As well as the highly reproducible material constants, the high current density after the contact point (high precision achieved through photolithographic manufacture) almost exclusively determines the sensor's resistance value. To eliminate polarity sensitivity the sensors are mostly series-connected in a pair in opposite alignment (double-hole version, Figure 5). The bottom electrode can then be designed as a metallic temperature contact (with no electrical function).

Measurement sensitivity is practically double that of a platinum resistor (TC 100 = $7.73 \cdot 10^{-3}$/K). However, the temperature-response curve is less linear than that of a metallic sensor. The measuring range is limited to approximately 150 °C by the material's intrinsic conductivity.

Applications in motor vehicles

These temperature sensors are used in many motor-vehicle applications. Here are some examples:

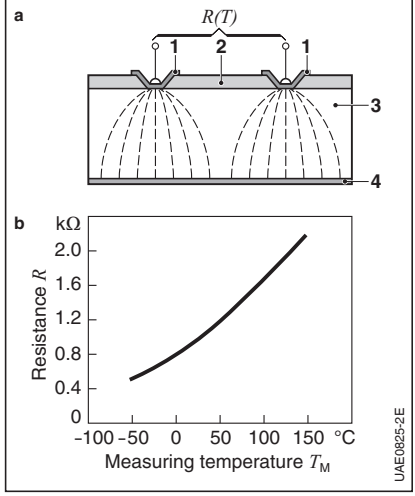
– Coolant-temperature sensor,
– Fuel-temperature sensor,
– Engine-oil temperature sensor,
– Intake-air temperature sensor,
– Charge-air temperature sensor,
– Outside-temperature sensor,
– Interior-temperature sensor,
– Exhaust-gas temperature sensor, etc.

References for temperature sensors

[1] DIN EN 60751: Industrial platinum resistance thermometers and platinum temperature sensors.

Figure 5: Silicon semiconductor resistor (spreading-resistance principle)

a) Design,
b) Characteristic curve.
1 Contacts,
2 Passivation (nitride, oxide),
3 Silicon substrate,
4 Counter-electrode without connection,
$R(T)$ Thermistor.



Torque sensor

Measured variables and application
Torque measurement is utilized in a variety of applications in motor vehicles. Torque sensors are used for example to sense the steering torque initiated by the driver.

Electromechanical power-steering systems are increasingly being used in vehicles. The essential benefits are simple installation and start-up in the vehicle, energy saving, and the suitability of these systems in the vehicle's ECU network for assistance systems to increase comfort and safety.

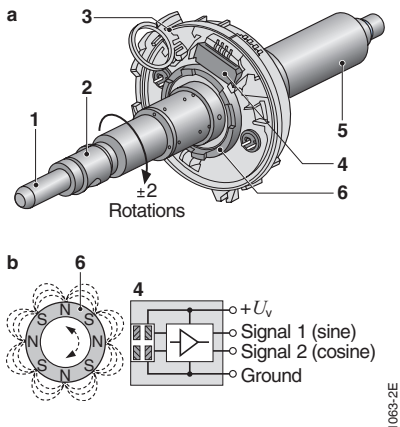
Measurement principle
When it comes to torque measurement, there are two different methods: angle and stress measurement. In contrast to stress-measurement methods (with strain gages), angle-measurement methods require a certain length of the torsion bar over which the torsion angle can be picked off.

To sense the driver command, it is necessary in an electromechanical power-steering system to measure the torque initiated by the driver. The sensors currently used in series production incorporate in the steering shaft a torsion bar which in the course of a steering torque applied by the driver undergoes a defined torsion linear to the initiated torque (Figure 1). The torsion is in turn measured by appropriate means and converted into electrical signals. The required measurement range of a torque sensor for use in an electromechanical power-steering system is usually around ±8 to ±10 Nm. The maximum torsion angle is mechanically limited by carrier elements to protect the torsion bar against overload and destruction.

To be able to measure the torsion and thus the applied torque, a magnetoresistive sensor is mounted on one side of the torsion bar which samples the field of a magnetic multipole wheel mounted on the other side. The number of poles on this wheel is chosen so that the sensor outputs a clear signal within its maximum measurement range to provide at all times a clear indication of the applied torque. The magnetoresistive sensor used supplies over the measurement range two signals which describe a sine signal and a cosine signal when shown against the torsion-bar torsion angle. The torsion angle and thus the torque are calculated in an ECU with the aid of an arc-tangent function.

Because the two signals are always permanently assigned over the defined measurement range, in the event of a deviation from these signals sensor faults can be detected and the necessary substitute measures initiated.

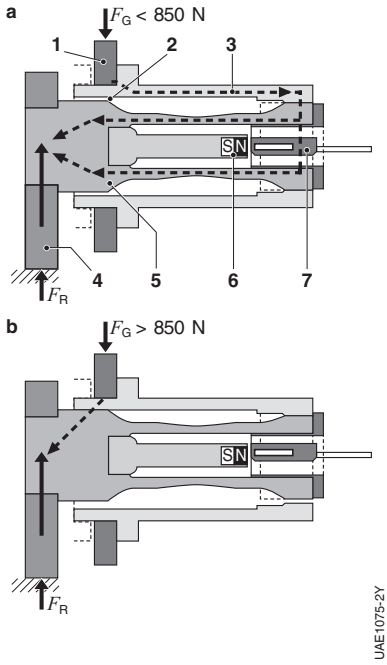
Figure 1: Torque sensor
a) Sensor module,
b) Measurement principle.
1 Torsion bar (torsion range internal),
2 Input shaft (from steering wheel),
3 Volute spring for electrical contacting,
4 Sensor module with magnetoresistive sensor chip and signal amplification,
5 Steering shaft,
6 Magnetic multipole wheel.



A volute spring with the required number of contacts is used to contact the sensor over the torsion range of approx. ±2 steering-wheel turns. This volute spring serves to deliver the supply voltage and transmit the measured values.

Figure 2: Measurement principle of iBolt force sensor

a) Conditions for F_G < 850 N (inside the measurement range),
b) Conditions for F_G > 850 N (outside the measurement range).
1 Swinging arm, 2 Air gap, 3 Sleeve,
4 Seat rail,
5 Double bending beam,
6 Magnet,
7 Hall IC.
F_G Weight,
F_R Bearing force.
The dashed lines indicate the flow of force in the force sensor.



Force sensor

Measured variables and application

One application of force measurement in motor vehicles is the sensing of the front passenger's weight. Classifying the front passenger by measuring their weight enables the airbag to be deactivated when a small child is sitting in the seat.

Measurement principle

The operating principle of the iBolt sensor is based on measuring the deflection of a bending beam in response to the front passenger's weight. The extent of this deflection is detected by measuring the magnetic field strength with a Hall sensor (Figure 2a).

The sensor is designed in such a way that preferably the vertical component of the front passenger's weights causes a deflection of the bending beam. The magnet and the Hall IC are arranged in the sensor in such a way that the static magnetic field that penetrates the Hall IC produces an electrical signal linear to the deflection of the bending beam. The special design of the sensor prevents a horizontal deflection of the Hall IC in respect of the magnet in order to minimize the influence of transverse forces and torques. In addition, the maximum stress in the bending beam is limited by a mechanical overload stop (Figure 2b). This protects the sensor particularly in overload situations in the event of a crash.

The force generated by the front passenger's weight is directed from the upper seat structure via the sleeve into the bending beam (Figure 2a). The force is then directed from the bending beam into the lower seat structure. The bending beam is designed as a double bending beam as this has an S-shaped deformation line. Here the two vertical connecting points of the double bending beam remain vertical for the entire deflection range. This ensures a linear and parallel movement of the Hall IC in respect of the magnet, producing a linear output signal.

Gas and concentration sensors

λ sensors

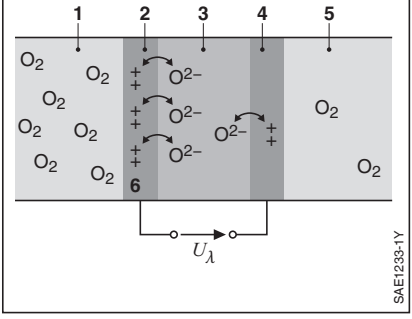
Measurement principle

λ sensors measure the oxygen content in the exhaust gas. They are used to regulate the air/fuel ratio in motor vehicles. The name is derived from the excess-air factor λ. This factor indicates the ratio of the current air volume to the theoretical air volume which is required for a full combustion of the fuel. It cannot be determined directly in the exhaust gas, but only indirectly via the oxygen volume present in the exhaust gas or required to fully convert combustible components. λ sensors consist of platinum electrodes which are mounted on a ceramic solid electrolyte that conducts oxygen ions (e.g. ZrO_2).

The signal from all the λ sensors is based on electrochemical processes with the involvement of oxygen (see Electrochemistry). The platinum electrodes used catalyze the reaction of remnants of oxidizable components in the exhaust gas (CO, H_2 and hydrocarbons $C_xH_yO_z$) with exhaust-gas oxygen. λ sensors consequently measure not the real oxygen content in the exhaust gas, but instead that content which corresponds to the chemical equilibrium of the exhaust gas.

Figure 1: Nernst cell

1 Reference gas, 2 Anode,
3 Solid electrolyte of Y-doped ZrO_2,
4 Cathode,
5 Exhaust gas,
6 Residual charges on the anode.
O^{2-} Oxygen ion,
U_λ Sensor voltage (Nernst voltage).



Applications

λ sensors are used in gasoline engines during the regulation of a stoichiometric mixture (λ = 1) to deliver the lowest-pollutant exhaust gas possible. Optimum impact by the three-way catalytic converter is guaranteed in this range.

Two-point λ sensors (step-change or switching-type sensors) indicate whether the mixture is rich (λ < 1, excess fuel) or lean (λ > 1, excess air). With these sensors the oxygen partial pressure of stoichiometric air/fuel mixtures can be measured very accurately due to the steep part of the sensor characteristic curve in this environment. Outside this environment, however, the characteristic curve is very flat (Figure 2).

Only the large measurement range of wide-band λ sensors (from λ = 0.6 to clean air) allow their use in systems with direct injection in stratified-charge operation and in diesel engines. λ control that can be achieved with wide-band λ sensors results in significant system and emission advantages over two-point control with a step-change sensor and further applications such as, for example, more accurate monitoring of the catalytic converter and controlled component protection. The high signal dynamic response of wide-band λ sensors with response times of less than 100 ms improves λ control.

Operating concept

All the exhaust-gas sensors described in the following consist of two modules: Nernst cell and pump cells.

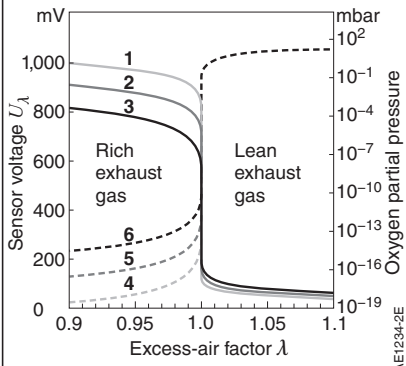
Nernst cell

The incorporation and removal of oxygen ions in the lattice of the solid electrolyte is dependent on the oxygen partial pressure on the surface of the electrode (Figure 1). Thus at low partial pressure more oxygen ions leave than enter. The vacant spaces in the lattice are reoccupied by advancing oxygen ions. An electric field is generated on account of the resulting charge separation at different oxygen partial pressures at the two electrodes. The electric field forces push advancing oxygen ions back and an equilibrium develops at the so-called Nernst voltage.

Pump cell
When a voltage is applied that is less or greater than the developing Nernst voltage, this state of equilibrium can be changed and oxygen ions actively transported through the ceramic. This creates between the electrodes a current, borne by oxygen ions. Crucial to the direction and intensity is the difference between the applied voltage (pump voltage U_P) and the developing Nernst voltage. This process is called electrochemical pumping.

Two-point λ sensors
Design
Two-point λ sensors indicate whether the mixture is rich (λ < 1, excess fuel) or lean (λ > 1, excess air). With these sensors the oxygen partial pressure of virtually stoichiometric air/fuel mixtures can be measured very accurately due to the steep part of the characteristic curve in this environment and the lowest-pollutant exhaust gas possible achieved by regulating the fuel quantity.

Figure 2: Characteristic curve of a two-point λ sensor at different sensor-element temperatures
U_λ Sensor voltage,
$p_A(O_2)$ Oxygen partial pressure in the exhaust gas,
λ Excess-air factor.
1 U_λ at 500 °C,
2 U_λ at 700 °C,
3 U_λ at 900 °C,
4 $p_A(O_2)$ at 500 °C,
5 $p_A(O_2)$ at 700 °C,
6 $p_A(O_2)$ at 900 °C.



Operating concept
The operating concept is based on the principle of a Nernst cell (Figure 1). The useful signal is the Nernst voltage U_λ that develops between an electrode exposed to the exhaust gas and an electrode exposed to the reference gas. The characteristic curve is very steep at λ = 1 (Figure 2). In lean mixtures the Nernst voltage increases linearly with the temperature. In rich mixtures, on the other hand, the influence of temperature on the oxygen partial pressure dominates in the state of equilibrium. The higher the temperature, the higher this oxygen partial pressure.

The establishment of equilibrium at the exhaust-gas electrode is also the cause of very small deviations of the λ jump from the exact value. The exhaust-gas electrode is covered with a porous protective ceramic layer to protect against contamination and to encourage the establishment of equilibrium by limiting the number of arriving gas particles. Hydrogen and oxygen diffuse through the porous protective layer and are converted on the electrode. More oxygen must be available on the protective layer to ensure that the faster-diffusing hydrogen is fully converted on the electrode; in all there must be a slightly lean mixture in the exhaust gas. The characteristic curve is therefore shifted in the lean direction. This "λ shift" is electronically compensated during control.

To deliver the signal a reference gas is required that is separated (gas-tight) from the exhaust gas by the ZrO_2 ceramic. Figure 3 depicts the structure of a planar sensor element with reference-air passage. Ambient air is used as reference gas in this type. Figure 4 shows the element in the sensor housing. The exhaust-gas and reference-gas sides are separated from each other (gas-tight) by means of a packing seal. The reference-gas side in the housing is permanently supplied with reference air via the supply leads.

Systems with a "pumped" reference are increasingly being used as an alternative to reference air. Pumping refers here to the active transportation of oxygen in the ZrO_2 ceramic by injecting a current, in which the current is set so low that it does not interfere with the actual measurement. The reference electrode itself is connected via a tighter output in the element to the reference-gas chamber. This results in a build-up of excess oxygen pressure at the reference electrode. This system offers additional protection against unwanted gas components penetrating into the reference-gas chamber.

Robustness

The ceramic sensor element is protected by a protective tube ahead of the direct exhaust-gas flow (Figure 4). This contains openings through which only a small amount of the exhaust gas is directed to the sensor element. It prevents heavy thermal stresses from the exhaust-gas flow and at the same time provides mechanical protection for the ceramic element.

The housing must meet stringent temperature requirements, meaning that high-quality materials must be used. Temperatures in excess of 1,000 °C can be recorded in the exhaust gas, as well as 700 °C at the hexagon head and up to 280 °C at the cable outlet. For this reason, only ceramic and metallic materials are used in the hot area of the sensor.

Most two-point λ sensors are also fitted with a heater element (Figure 3). This quickly heats (FLO, fast light-off) the sensor element to operating temperature and provides for fast control readiness.

Figure 4: Two-point λ sensor, sensor element in the housing

1 Protective tube,
2 Sensor element,
3 Hexagon head,
4 Reference gas,
5 Electrical supply lead,
6 Exhaust-gas side,
7 Packing seal,
8 Support ceramic,
9 Contacting.

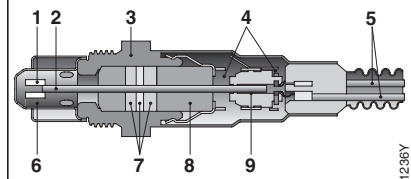
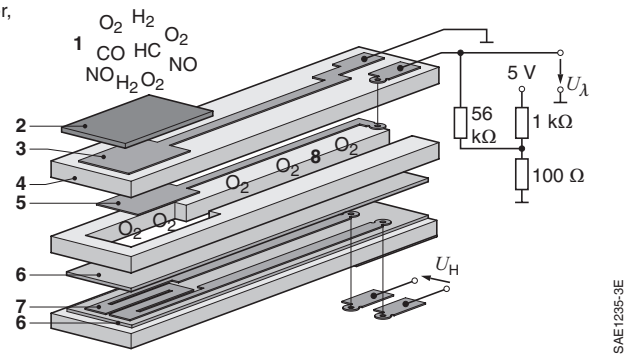


Figure 3: Structure of a planar two-point λ sensor with protective circuit (exploded view)

The vertical lines symbolize conductive connections.

1 Exhaust gas,
2 Porous protective layer,
3 Outer electrode (platinum electrode),
4 ZrO_2 ceramic with Nernst cell,
5 Reference electrode (platinum electrode),
6 Al_2O_3 insulation layer,
7 Heater element,
8 Reference air.
U_λ Sensor voltage,
U_H Heating voltage.



In practice the λ sensor after engine starting is switched on only after a delay. Water that is produced as a combustion product and condenses again in the cold exhaust-gas system is transported by the exhaust gas and can reach the sensor element. If a droplet of this water touches a hot sensor element, it vaporizes immediately and abstracts a large amount of heat locally from the sensor element. The heavy mechanical stresses that occur as a result of thermal shock can cause the ceramic sensor element to break. For this reason, the sensor is often switched on only after the exhaust-gas system has been sufficiently heated. In more recent developments the ceramic elements are enclosed in a further porous, ceramic layer which significantly increases robustness against thermal shock. When a water droplet touches the elements, it spreads in the porous layer. Local cooling is spread over a wider area and mechanical stresses are reduced.

Protective circuit

Figure 3 shows the protective circuit of a two-point λ sensor. Because the sensor when cold cannot generate a signal due to the absence of conductivity of the ZrO_2 ceramic, it is connected via a resistor to a voltage divider. In the cold state the sensor signal is therefore 450 mV, the value of a stoichiometrically combusted gas ($\lambda = 1$). As the temperature increases the sensor is able to develop the Nernst voltage. After approximately 10 s the sensor is at sufficiently high temperature to indicate externally specified lean/rich changes. It is then possible to switch to closed-loop control in the vehicle.

Types

Two-point λ sensors come in different types. The sensor elements can be shaped in the form of a finger with a separate heater element or as a planar element with an integrated heater element which is manufactured using film technology (Figure 3).

Wide-band λ sensor

Design and function

With the two-point λ sensor the oxygen partial pressure of stoichiometric air/fuel mixtures can be measured very accurately in the steep part of the characteristic curve. However, where there is excess air ($\lambda > 1$) or excess fuel ($\lambda < 1$), the characteristic curve is very flat (Figure 2).

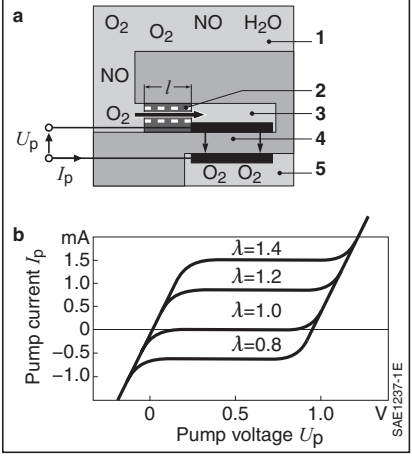
The large measurement range of wide-band λ sensors ($0.6 < \lambda < \infty$) enables them to be used in systems with direct injection and stratified-charge operation and in diesel engines. A continuous control concept that can be achieved with a wide-band λ sensor delivers significant system advantages, such as for example controlled component protection. The high signal dynamic response of wide-band λ sensors ($t_{63} < 100$ ms) improves exhaust emissions to the point of low-emission vehicles, which necessarily calls for measures such as individual-cylinder closed-loop control.

Design and operating principle

The Nernst principle is reversed to wide the measurement range. When a voltage greater than the Nernst voltage is applied to the measuring cell, oxygen is transported as oxygen ions through the pump-cell ceramic to the internal pump and discharged there as free, molecular oxygen into the reference-air chamber. The resulting pump voltage is obtained from the difference between pump voltage (U_P) and the developing Nernst voltage.

The inflow of oxygen molecules is limited in order to establish a linear connection with the externally applied oxygen concentration. To this end the admission of exhaust-gas is limited by a porous, ceramic structure with specifically set pore radii, the so-called diffusion barrier.

Figure 5: Once-cell λ sensor in lean exhaust gas
a) Cross-section,
b) Characteristic curves.
l Length of diffusion barrier.
1 Lean exhaust gas,
2 Diffusion barrier,
3 Cavity,
4 Pump cell,
5 Reference air.
The arrows in the pump cell indicate the pumping direction.



The pump current I_P that flows at sufficient pump voltage is, on account of the law of diffusion, directly proportional to the partial pressure in the exhaust gas:

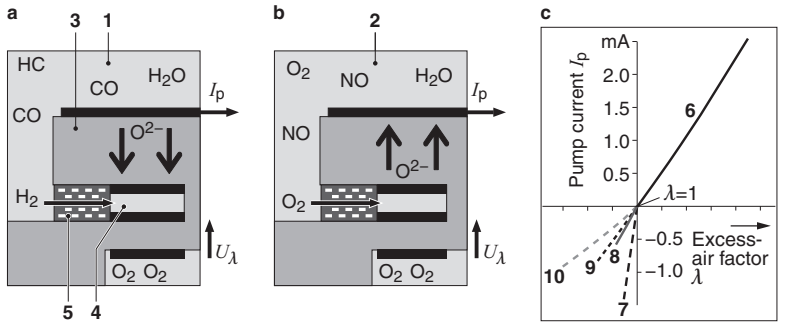
$$\frac{I_P}{4F} = I_M = \frac{A\,D(T)}{R\,T\,l}\left(p_A(O_2) - p_H(O_2)\right).$$

Here $p(O_2)$ is the oxygen partial pressure in the exhaust gas ($p_A(O_2)$) or in the cavity ($p_H(O_2)$), T the temperature, $D(T)$ the temperature-dependent diffusion constant of the diffusion barrier, A its cross-sectional area, and l its length (Figure 5). F is the Faraday constant (F = 96485.3365 C/mol), R the general gas constant (R = 8.3144621 J/mol·K).

If rich exhaust gas is present, the created Nernst voltage of approx. 1,000 mV counteracts the applied pump voltage such that the resulting negative voltage pumps oxygen in the opposite direction into the cavity and thus the linear shape of the characteristic curve is extended into the rich range. The oxygen is obtained for this purpose from the reduction of water and CO_2 on the outer electrode.

The drawback of this simple form of wide-band λ sensor is that the fixed pump voltage must be sufficient in rich exhaust gas to pump oxygen into the cavity and in

Figure 6: Two-cell λ sensor in rich and lean exhaust gas
a) and b) Cross-section,
c) Sensor characteristic curves.
Depending on the polarity of pump current I_p predominantly reducing exhaust-gas constituents (section a) or oxygen (section b) diffuse through the diffusion barrier.
For $\lambda < 1$ the characteristic curve (section c) is dependent on the exhaust-gas composition, here the characteristic curves of individual exhaust-gas constituents are plotted.
1 Rich exhaust gas, 2 Lean exhaust gas, 3 Pump cell, 4 Nernst cell, 5 Diffusion barrier,
6 Curve for O_2, 7 Curve for H_2, 8 Curve for CH_4,
9 Curve for CO, 10 Curve for C_3H_6.



lean exhaust gas to pump oxygen out of it. The internal resistance of the pump cell should therefore be very low. In addition, the measurement range in the rich range is limited by the molecule inflow in the reference passage. An oxygen deficiency in the reference caused by a reference-air contamination or over-range cannot be clearly distinguished. In the case of exhaust-gas exchanges the dynamic response of the one-cell sensor is limited by recharging of the electrode capacities when the pump voltage is changed.

To eliminate these drawbacks, the Nernst cell familiar from the step-change sensor is combined with the oxygen pump cell described above. In the case of rich exhaust gas (Figure 6a), in response to reversing the voltage at the outer pump electrode, oxygen is generated from H_2O and CO_2, transported inwards through the ceramic, and discharged again in the cavity. There the oxygen reacts with the diffusing, reducing exhaust-gas components. The created inert reaction products H_2O and CO_2 diffused through the diffusion barrier in an outward direction.

The oxygen partial pressure in the cavity is measured by the Nernst cell also connected to the cavity and the pump voltage is corrected by a controller (Figure 7) to a specified reference voltage (e.g. 450 mV). There is then an oxygen partial pressure of 10^{-2} Pa in the cavity.

Because the diffusion limit current increases with the sensor temperature, the temperature must be kept as constant as possible. The highly temperature-dependent resistance of the Nernst cell is measured for this purpose. The operating electronics regulate it by means of the heater element operated with a pulse-width-modulated voltage.

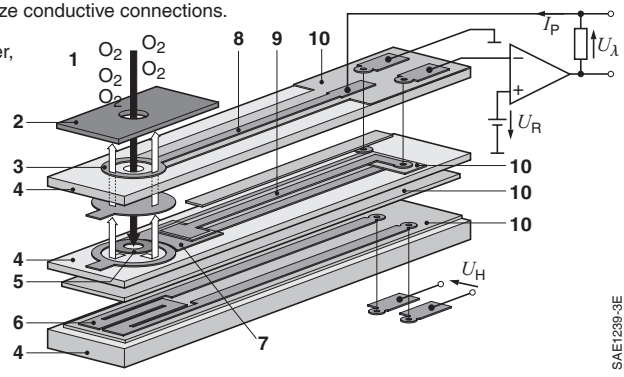
Where the exhaust gas is rich the different diffusion coefficients, which are functions of the masses of the gas molecules, become noticeable through different sensitivities (Figure 6). The signal is therefore used in the ECU for a characteristic curve applied for the relevant gas composition.

The sensor's diffusion limit current and thus the sensitivity are very much dependent on the geometry of the diffusion barrier. It is necessary to adjust the pump current in order to achieve the high accuracy requirement in the manufacturing tolerance. In some sensor types this is done by a hybrid resistor in the sensor plug which acts together with the measuring resistor as current divider. Alternatively, the diffusion limit current can be adjusted already in the sensor-element production process by means of specific openings so that there is no need to make an adjustment at the plug. For subsequent calibration of the sensor in the vehicle the oxygen concentration of the air can be measured in overrun conditions and thus the characteristic curve corrected in the ECU.

Figure 7: Exploded view of a wide-band λ sensor

The vertical lines symbolize conductive connections.

1 Exhaust gas,
2 Porous protective layer,
3 Pump cell,
4 ZrO_2 ceramic,
5 Diffusion barrier,
6 Heater element,
7 Nernst cell,
8 Pump electrode,
9 Reference electrode,
10 Al_2O_3 insulation.
I_p Pump current,
U_R Reference voltage,
U_λ Sensor voltage,
U_H Heating voltage.



NO_x sensor

Application

NO_x sensors are used in the denitrification systems of diesel and gasoline engines. In systems with diesel engines they are installed upstream and downstream of SCR catalytic converters (selective catalytic reduction) and downstream of NO_x storage catalytic converters (NSC, NO_x storage catalysts). In systems with gasoline engines they are used downstream of the NO_x storage catalytic converter.

In these positions the NO_x sensors determine the nitrogen-oxide and oxygen concentrations in the exhaust gas and downstream of SCR catalytic converters also the ammonia concentration as a cumulative signal. In this way the engine-management system receives the current residual concentration of nitrogen oxides, provides for exact metering, and detects any faults in the exhaust system.

Nitrogen oxides are stored as nitrates in NO_x storage catalytic converters. In short rich phases the catalytic converter is regenerated, whereby the nitrates are reduced with the aid of carbon monoxide and hydrogen to nitrogen.

Design and operating principle

The NO_x sensor in Figure 8 is a planar three-cell limit-current sensor. A Nernst concentration cell and two modified oxygen pump cells (oxygen pump cell and NO_x cell), as are already familiar from wide-band λ sensors, form the overall sensor system.

The sensor element consists of several oxygen-conducting, ceramic solid-electrolyte layers insulated against each other, on which six electrodes are mounted. The sensor is fitted with an integrated heater element which heats the ceramic to an operating temperature of 600 °C to 800 °C.

The outer pump electrode exposed to the exhaust gas and the inner pump electrode in the first cavity, which is separated from the exhaust gas by a diffusion barrier, form the oxygen pump cell.

Figure 9: Characteristic curve of the nitrogen-oxide signal

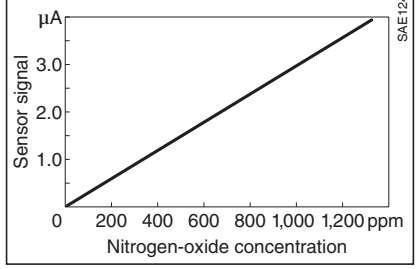
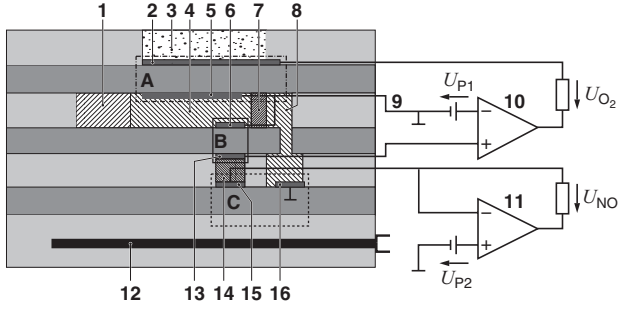


Figure 8: Cross-section of a NO_x sensor

A Oxygen pump cell, B Nernst cell, C NO_x cell.
1 Diffusion barrier 1, 2 Outer pump electrode, 3 Porous aluminum-oxide layer,
4 First cavity, 5 Inner pump electrode, 6 Nernst electrode, 7 Diffusion barrier 2,
8 Second cavity, 9 Common return conductor, 10 Oxygen regulator with voltage transformer,
11 NO current amplifier and voltage transformer, 12 Heater element, 13 Reference electrode,
14 Reference-gas chamber, 15 NO_x counter-electrode, 16 NO_x pump electrode.



The first cavity contains the Nernst electrode, while the reference-gas chamber contains the reference electrode. Together they form the Nernst cell. These are the functional components which are identical to those of wide-band λ sensors.

In addition, there is a third cell – the NO_x pump electrode and its counter-electrode. The former is situated in a second cavity, which is separated from the first by a further diffusion barrier, the latter is located in the reference-gas chamber. All the electrodes in the first and second cavities have a common return conductor.

The inner pump electrode is, unlike the inner pump electrode in wide-band λ sensors, greatly limited in its catalytic activity by the alloying of platinum with gold. The applied pump voltage U_{P1} is only enough to split (dissociate) oxygen molecules. NO is only minimally dissociated at the adjusted pump voltage and passes the first cavity with low losses. NO_2 as a strong oxidizing agent is converted on the inner pump electrode directly into NO. Ammonia reacts to this in the presence of oxygen and at temperatures of 650 °C into NO and water.

Because of the higher voltage at the NO pump electrode and its catalytically improved activity due to the admixture of rhodium, NO is completely dissociated at this electrode and the oxygen is pumped away through the solid electrolyte.

Electronics

In contrast to the other ceramic exhaust-gas sensors, the NO_x sensor is equipped with evaluation electronics (SCU, sensor control unit). It delivers via CAN bus the oxygen signal, the NO_x signal, and in each case the status of these signals.

Incorporated in the evaluation electronics are a microcontroller, an ASIC (Application-Specific Integrated Circuit) for operating the oxygen pump cell, and a high-precision instrument amplifier for the very small NO signal currents.

Characteristic curves

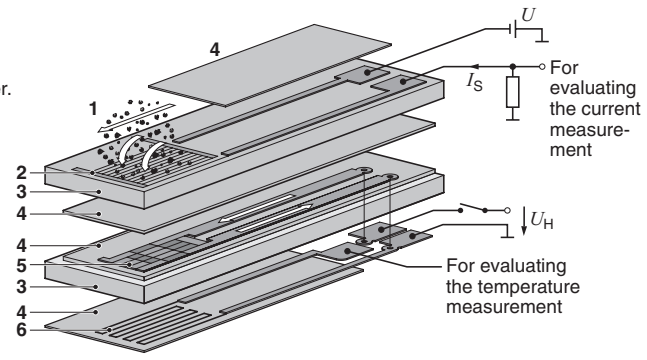
The oxygen signal is 3.7 mA for air. The oxygen characteristic curve is identical to that familiar from wide-band λ sensors (Figure 6c). The NO_x characteristic curve is shown in Figure 9.

Particulate sensor

The emission regulations in the USA and in Europe necessitate the use of diesel particulate filters (DPF) in diesel vehicles. Particulate sensors in addition to the differential pressure sensor are required to comply with future on-board diagnosis legislation (OBD), which places more stringent demands on the monitoring of the functional capability of these particulate filters. The particulate sensors measure the soot emissions downstream of the particulate filter.

Figure 10: Exploded view of the particulate sensor

1 Exhaust gas,
2 Interdigital structure with two comb-shaped electrodes,
3 Ceramic,
4 Insulation,
5 Heater element,
6 Platinum meander.
U Direct voltage, e.g. 60 V,
I_S Sensor current,
U_H Heating voltage.



Design
The sensor element consists of an interdigital structure with two comb-shaped platinum electrodes on a ceramic substrate with an integrated heater element, as is familiar from the λ sensor (Figure 10). A platinum meander serves to measure the temperature of the sensor element.

Operating principle
The particulate sensor is a resistive sensor. A direct voltage of for example 60 V is applied at the interdigital structure with initially very high electrical resistance. As a result of the field forces soot particles from the exhaust gas collect on the interdigital electrodes and form increasingly conductive soot paths between the two comb-shaped structures. This produces a monotonously increasing current between the electrodes (Figure 11). A predefined current threshold value is reached after a certain collection time. This triggers regeneration, in which the initial state is re-established when the sensor element is heated and the soot is burned off at temperatures above 600 °C.

The time between the start of measurement and the start of regeneration is defined as the trigger time. An integrated temperature-measurement meander is used to control the temperature to ensure that a controlled regeneration of the sensor element takes place.

The particulate-sensor element is, like that of the λ sensor, installed in a sensor housing. Because particulate deposits on the sensor element are desired here, particular emphasis is placed on the design of the protective tube.

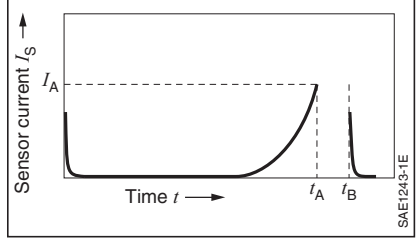
Electronics
The particulate sensor, like the NO_x sensor, is equipped with evaluation electronics. The latter supply the sensor current and the signal status via a CAN bus. A microcontroller controls the timing of measurement and regeneration and the compensation of data in response to the measured sensor-element temperature. The electronics also include a voltage stabilizer, a CAN driver, and the output stage for the heater element.

Hydrogen sensors
The use of fuel cells in motor vehicles necessitates the incorporation of hydrogen sensors. This sensor system has two functions – safety monitoring with leak detection (measurement range 0 to 4 % hydrogen content in air) and process control with adjustment of operating conditions (measurement range 0 to 100 %).

The most common measurement principles are described in descending frequency of use. For the most part different methods are combined to reduce cross-sensitivity to other gases or to increase sensitivity and widen the measurement range [3].

Electrochemical measurement principle
In the case of amperometric measurement a constant voltage is applied to an electrolyte that conducts only protons (e.g. sulfonated tetrafluoroethylene polymer). The electrolyte is covered by a diffusion barrier which is very selective in admitting only hydrogen. These sensors thus have long-term stability. At the electrode hydrogen splits into protons, which flow as pump current through the electrolyte. The electrons flow through the electrodes, driven by the applied voltage, via the outer electric circuit to the other side of the electrolyte. Here they recombine with oxygen on the electrode to form water. The current intensity is according to Faraday's law proportional to the number of protons, which in turn is determined by the diffusion by the diffusion barrier (see λ sensor).

Figure 11: Current curve between the electrodes of a particulate sensor
I_A Trigger threshold,
t_A Trigger point for regeneration,
t_B Start of next measurement cycle.

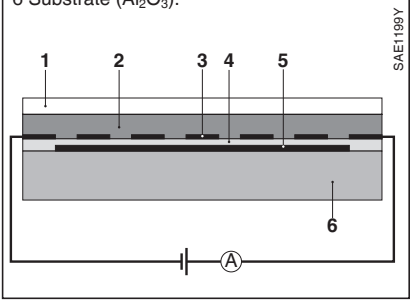


Potentiometric measurement methods utilize the potential difference between two electrodes. One electrode with a platinum or palladium catalyst is situated in the hydrogen environment, the other electrode is situated for example in air. According to the Nernst equation a potential difference develops (see Fuel cell). The potentiometric sensors are cross-sensitive to carbon monoxide that covers the catalyst. The age quickly and have a high drift. The measurement range is between 100 ppm and 100 %.

Resistance-based measurement principle
Palladium and semiconductor metal oxides (e.g. SnO_2) exhibit a change in resistance after the adsorption of hydrogen. The semiconductor metal-oxide sensor (Figure 12) consists of a gas-filtering diaphragm which should only admit hydrogen. The sensitive layer absorbs the hydrogen and induces the change in resistance, which is measured by a Wheatstone bridge.

A MOSFET (see Field-effect transistor) can also be used with a palladium gate as a variant. These alters the current between source and drain proportionally to the hydrogen absorbed at the gate. In a Schottky-diode arrangement (see Schottky diode) the breakdown voltage decreases when hydrogen is adsorbed.

Figure 12: Resistance-based measurement principle of a hydrogen sensor
1 Gas-filtering diaphragm,
2 Sensitive layer,
3 Semiconductor metal-oxide film,
4 Insulating layer,
5 Heater,
6 Substrate (Al_2O_3).



The measurement range for the resistance-based measurement principles is between 10 ppm and 2 %.

Catalytic measurement principle
Heat is generated when adsorbed gases oxidize on the surface of a catalyst. In a pellistor a wire consisting of catalyst material (e.g. platinum or palladium) exposed to the hydrogen is heated and as a result undergoes a change in resistance dependent on the hydrogen concentration, while the temperature and resistance of a reference wire remain unchanged. The resistors are integral parts of a Wheatstone-bridge circuit, with which the change in resistance is measured and from this the hydrogen concentration is inferred.

In the thermoelectric method the voltage build-up due to the temperature difference (Seebeck effect) is utilized.

The catalytic methods are cross-sensitive to other oxidizable gases and require a minimum quantity of oxygen (5 to 10 %) as the oxidation partner; the measurement range therefore is between 1 % and 90 to 95 % hydrogen concentration.

References for gas and concentration sensors
[1] Nernstgleichung: Zeitschrift für physikalische Chemie, IV. Volume Book 1, Verlag Wilhelm Engelmann, 1889, Published by W. Ostwald, J.H. Van't Hoff, W. Nernst: Die elektromotorische Wirksamkeit der Ionen.
[2] T. Baunach, K. Schänzlin, L. Diehl. Sauberes Abgas durch Keramiksensoren. Physik Journal 5 (2006) No. 5.
[3] T. Hübert et al.: Hydrogen sensors – A review, Sensors and Actuators B 157 (2011) 329 – 352.

Optoelectronic sensors

Internal photoelectric effect

The internal photoelectric effect forms the basis of optoelectronic sensor elements. Light can be seen as a stream of individual light quanta (photons). The energy E_{Ph} of a photon is only dependent on its frequency f or its wavelength λ:

(1) $E_{Ph} = h f = \frac{h c}{\lambda}$

where
h Planck's action quantum,
c Speed of light.

When photons strike atoms, they can each release an electron from the outer electron shell if the energy is sufficient. The energy required for release corresponds to the difference between the energy level E_V of the valence band of the atom and the level E_L of the conduction band, i.e. the band gap E_g.

(2) $E_g = E_L - E_V$.

To release an electron, the photon energy E_{Ph} must therefore be greater than the band gap E_g. In a pure semiconductor, charge-carrier pairs (electrons and holes) are created by the absorption of photons. The band gap to be overcome of, for example, silicon is E_g = 1.12 eV at room temperature. Without special measures the charge-carrier pairs created recombine already after a short time. However, the radiation created in the process is not in the visible range in the case of silicon.

In highly doped semiconductors, the above-mentioned intrinsic photoelectric effect is supplemented by the extrinsic photoelectric effect. Because the energy gap to be overcome is significantly lower in such extrinsic sensors, they are also suitable for radiating a larger wavelength (infrared range).

No further release occurs for energies $E_{Ph} < E_g$. According to equation (1), this corresponds in the case of silicon to a limit wavelength of λ_g = 1.1 µm (near infrared). Light with larger wavelengths or lower frequency is not longer absorbed; silicon becomes transparent here.

Light-sensitive sensor elements

Photoresistors

Incident light causes charge-carrier pairs to be created in a sensor configured as a resistor (LDR, light-dependent resistor); these charge-carrier pairs increase the conductance G. They recombine after a short time (milliseconds range); but nevertheless the charge-carrier concentration increases in stationary equilibrium with the illuminance E according to the following law:

(3) $G = \text{const} \cdot E^{\gamma}$ where γ = 0.7 to 1.

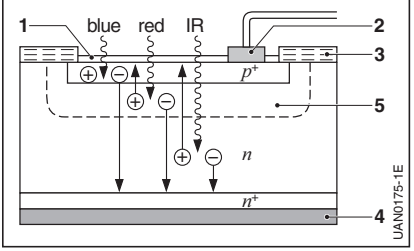
The light-sensitive materials usually used are cadmium sulfide CdS (E_g = 1.8 eV; λ_g = 0.7 µm) and cadmium selenide CdSe (E_g = 1.5 eV; λ_g = 0.8 µm) on ceramic substrates.

Semiconductor pn junctions

There is essentially no difference between a photocell, a photodiode, and a phototransistor. They all use the photocurrent or no-load voltage in illuminated *pn* semiconductor contacts as the measuring effect. The elements differ, however, in the way they operate.

Charge carriers which are created by the internal photoeffect in the depletion layer of a *pn* semiconductor contact (Figure 1) are immediately accelerated by the electric field in the space-charge region with low charge-carrier concen-

Figure 1: Separation of created electron-hole pairs in a planar semiconductor component with *pn* junction
1 Optical coating,
2 Contact,
3 SiO_2,
4 Metal contact,
5 Space-charge region.



tration, as a result of which the charge carriers are immediately split after they have been created (drift current). Their recombination is practically prevented and photosensitivity is significantly increased.

Photocells

Photocells are operated without external bias and can be operated both at no load (photovoltaic effect) and in a short-circuit. They have accordingly a low background noise and thus a high detection capability.

The characteristic curves applicable to these operating modes (Figure 2) can be easily derived as special cases from a diode polarized in the forward direction with the voltage U with the thermally conditioned off-state saturation current I_S and the photocurrent I_{ph} also flowing in reverse direction:

(4) $I = I_S\, e^{\frac{eU}{kT}} - I_S - I_{ph}$

where
e Elementary charge,
k Boltzmann's constant,
T Absolute temperature.

Special cases:

(5) $U = 0$ (short-circuit)
$\rightarrow I = I_K = -I_{ph}$;

(6) $I = 0$ (no load):

$$\rightarrow U = U_L = \frac{kT}{e} \cdot \ln\left(\frac{I_{ph}}{I_S} + 1\right).$$

Photocells are usually configured with a very large irradiation-sensitive surface, and provide accordingly relatively high photovoltaic current also (e.g. I_{ph} = 250 µA at E = 1,000 lx). Their time constant is proportionally high and is typically approximately 20 ms.

Photodiodes and phototransistors

Photodiodes are operated with constant bias U_S in the reverse direction, where the photocurrent flowing as reverse current is linearly dependent on the illuminance E (Figure 3). The space-charge region is increased in size by the applied

Figure 3: Characteristic curves for a photodiode for constant illuminance *E*

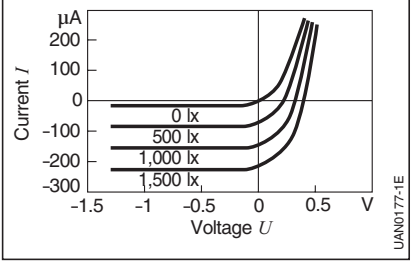


Figure 2: Characteristic curves for a photocell as a function of illuminance *E*

a) Short-circuit current I_K,
b) Open-circuit voltage U_L.

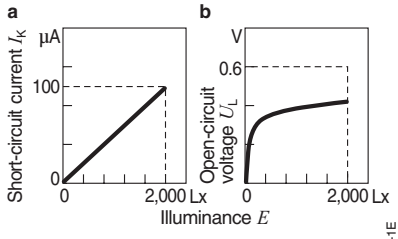
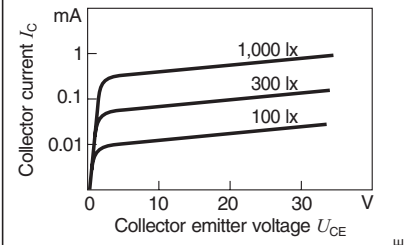


Figure 4: Characteristic curves for a phototransistor for constant illuminance *E*



reverse voltage. This reduces the junction capacitance in such a way that the cutoff frequency of such a photodiode is typically a few MHz.

In the case of the phototransistor shown in Figure 4 (*npn* type), the collector-base diode polarized in the reverse direction acts as a photodiode. In this way, the collector, like every transistor, supplies a photocurrent higher by the current-amplification factor B (≈ 100 to 500) (corresponding to base current). However, the higher sensitivity comes at the expense of a slightly worse frequency dynamic and a slightly worse thermal characteristic.

Application

Dirt sensor

The sensor measures the level of contamination on the headlamp lens to furnish the data required for automatic lens cleaning systems.

The sensor's photoelectric reflected-light barrier consists of a light source (LED) and a light receiver (phototransistor). The source is positioned on the inside of the lens, within the cleansed area (Figure 5), but not directly in the headlamp beam path. When the lens is clean, or covered with rain droplets, the infrared measurement beam emitted by the unit passes through the lens without being obstructed. Only a minuscule part is reflected back to the light receiver. However, if it encounters dirt particles on the outer surface of the lens, it is reflected back to the receiver at an intensity that is proportional to the degree of contamination and automatically activates the headlamp washer unit once a defined level is reached.

Rain sensor

The rain sensor recognizes rain droplets on the windshield, so that the windshield wipers can be triggered automatically. The driver, however, can still use the manual controls.

The sensor consists of an optical transmission and reception path (similar to the dirt sensor). However, the light is directed toward the windshield at an angle. A dry outer surface reflects all the light (total reflection) back to the receiver which is also mounted at an angle (Figure 6). When water droplets are present on the outer surface, a substantial amount of the light is refracted outward, thus weakening the return signal. The windshield wiper also responds to dirt once the activation threshold is exceeded.

Figure 5: Dirt sensor for headlamp lens
1 Lens,
2 Dirt particle,
3 Sensor element,
4 Transmitter,
5 Receiver.

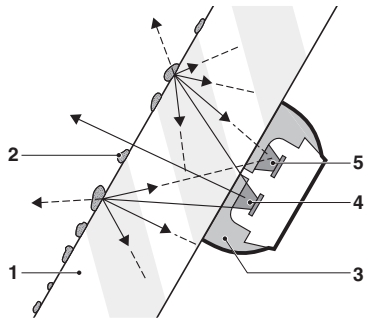
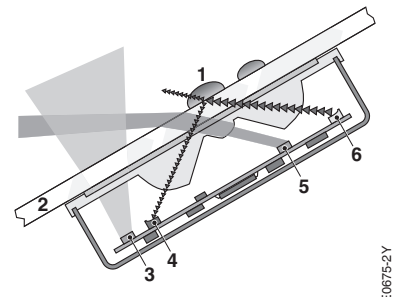


Figure 6: Rain sensor for windshields
1 Raindrop,
2 Windshield,
3 Ambient-light sensor,
4 Photodiode,
5 Light sensor aligned to far distance,
6 LED.



Infotainment and cockpit solutions

History

In-car entertainment had already been available in the 1920s with the first car radio in the legendary Ford Model T. In 1933 the Berlin-based Radiotelefon und Apparatefabrik Ideal AG, a subsidiary of Bosch, brought out a car radio for the medium- and long-wave ranges – VHF reception first became available only in 1949 – with five electron tubes. This appliance was developed together with Bosch. Station tuning was at that time performed with Bowden cables from the steering column; only later were all the components integrated in the instrument panel. The volume comprised around ten liters.

Enhanced comfort arrived in the 1950s with automatic station search and push-buttons for storing several stations. Appliances became more compact with the introduction of transistor technology.

In 1968 the first car radios with integrated cassette players came out. This made individual in-car entertainment possible.

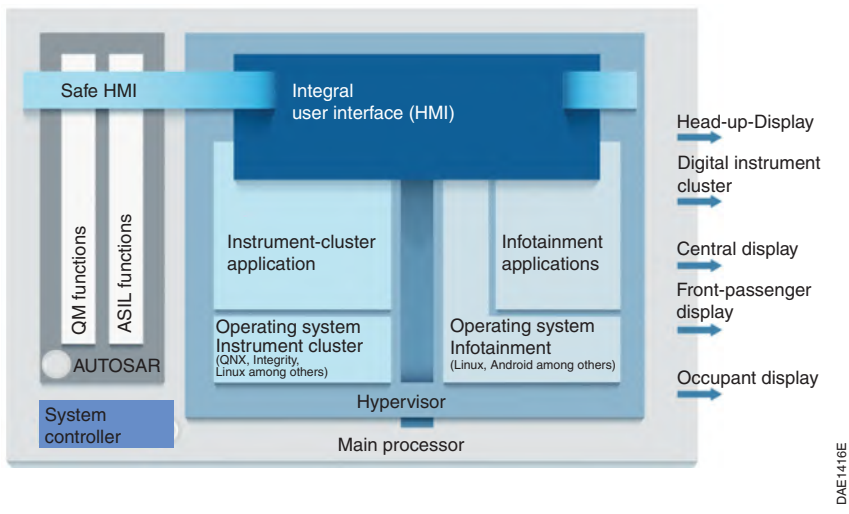
In 1974 the Autofahrer-Rundfunk-Informationssystem (ARI), a radio traffic-information system for drivers, was launched in Germany. This enabled traffic information to be automatically detected on the car radio.

In the mid-1980s the first car radios with an integrated CD player were introduced, delivering bettering sound quality than the cassette player.

In 1988 the Radio Data System (RDS) was launched. This enabled a car radio to automatically process traffic reports and receive station names and further additional information. Later, with the Traffic Message Channel (TMC), the RDS traffic reports could also be shown independently of the received radio program in the car radio's display.

In the 1990s navigation devices found their way into cars, giving rise to combinations of car radio, CD player, and navigation systems.

Figure 1: Information Domain Computer



Infotainment systems

Technology of the present day

Entertainment and information have played a significant role in motor vehicles for a long time. The demands imposed on the units have increased continuously.

Quickly updating the favorites playlist before starting a journey and checking the current traffic situation en route in real time – infotainment systems have long been more than just a radio. Integrated in these systems are entertainment components, intelligent vehicle navigation, telematics, and driver-assistance systems. Connectivity with the Internet and the vehicle bus system facilitates communication between driver, vehicle, and outside world. Intuitive operation minimizes driver distraction, making driving safer. The HMI (Human Machine Interface) as the interface between human and machine is optimized here for distraction-free driving.

Such systems in the meantime have become standard features in motor vehicles. The car radio and the navigation system are integrated in the central display in the center console and controlled via the touchscreen. Speedometer, rev counter, coolant temperature, and further information such as for example navigation-system route-direction symbols are displayed in the instrument cluster.

With a fully digital instrument cluster the electromechanical indicating instruments for engine speed, torque and coolant temperature are rendered superfluous. The indicating instruments are reproduced electronically.

In the additional front-passenger display the front passenger can for example watch a video or surf the Internet without the driver being distracted.

Solutions for the future

The demands imposed on the vehicle cockpits of the future are high. As with the smartphone or other digital companions, car drivers expect the greatest possible connectivity, personalization and intuitive operation from their vehicles. Vehicle manufacturers must offer suitable solutions for different markets and vehicle segments and in the process expect the greatest possible flexibility with low costs.

Information Domain Computer

System overview

The "Information Domain Computer" from Bosch (Figure 1) is based on a modular platform concept and will fulfill all these requirements. It provides a consistent, networked user experience, satisfies the high safety requirements, and offers car makers a high degree of flexibility in the configuration of their solution.

At present, multiple ECUs control and regulate the different displays and other electronic cockpit functions in production vehicles. In future more and more functions will be merged in one single processor. The Information Domain Computer is based on a "System on a Chip" (SoC) with the main processor controlling all the functions on just one chip. All the displays in the vehicle can be activated by this Domain Computer. This reduces the complexity and the demand for additional ECUs in the cockpit. The computing functions of the previously separate domains of infotainment and Instrumentation as well as other functions are bundled on one processor. This saves costs, installation space, and weight and consumes less energy.

An ECU for the entire user interface – the merging of the previously separate systems in the cockpit facilitates the synchronization of the infotainment system, the instrument cluster and other displays so that all the given information can be orchestrated, managed and displayed at any time in the entire vehicle. This is a prerequisite for a seamless, uniform design language and the possibility of always flexibly displaying information where the driver needs it. The complete system is updatable "over-the-air", i.e., the Information Domain Computer can be updated in real time.

The infotainment system integrates numerous functions such as the media player, radio reception, intelligent navigation, smartphone integration, voice recognition, and third-party software.

The hypervisor enables the different operating systems for example for infotainment and driver-information displays to be operated in separate virtual machines.

Safety-relevant functions from the domains of infotainment and instrument

cluster run on the AUTOSAR operating system. Examples are the rearview camera and an oil-level warning display, which must be immediately reliably available when the vehicle is started.

The system controller has a shorter boot time than the main processor itself and thus guarantees the immediate availability of all the safety-relevant functions.

Digital instrument cluster
As well as the classic instrument-cluster displays, further functions such as climate control, driver-assistance functions, and driver-monitoring camera with their own operating systems can be optionally integrated. Thanks to the intelligent algorithms only such information that the driver needs depending on the situation is displayed in the digital instrument cluster. This reduces the risk of the driver being overwhelmed by the increasing amount of information from driver-assistance functions. Displayed content can be flexibly adapted, reducing the degree of driver distraction and at the same time helping to improve road safety.

Because a lot of driver information is safety-relevant, modern digital instrument clusters must offer adaptive concepts (e.g., the “Safe HMI” for satisfying ASIL requirements). Complex functionalities and technologies, such as for example the driver-monitoring camera, can also be integrated in the instrument cluster.

Smartphone connection to the infotainment system

Application
Safely, comfortably and without distraction – mySPIN is a platform-independent solution from Bosch that is designed to let the user connect their smartphone easily to their car or motorcycle. In this way, selected apps and networked services are displayed in a driver- and rider-friendly format in the vehicle's cluster.

Once the smartphone is connected to the vehicle via WLAN, Bluetooth or USB, many interesting apps and important smartphone functions (e.g., telephony) are available to the driver or rider in the vehicle display. The user has access to their smartphone at all times during their journey. mySPIN offers comfortable use of apps both in passenger cars and on motorcycles or two-wheelers of any kind.

Whether iOS or Android, once connected to the vehicle, the driver or rider can use compatible smartphone apps directly via the touchscreen – on a motorcycle quite easily via the handlebar control unit. So as not to distract the driver or rider from the surrounding traffic and to provide driver- and rider-friendly and safe operation, all apps have been specially adapted to mobile use and exclusively make available all the important basic functions and information.

mySPIN already today offers the user an impressive selection of popular apps from a wide variety of categories, including navigation, point of interest and events, parking search, news, and music streaming. The pool of mySPIN-compatible apps that provide international as well as local content for regions like Europe, America, India, and China is growing continuously.

mySPIN offers the automobile manufacturers a great many interesting options too, including a seamless integration of the application in the vehicle's user interface and the facility to process vehicle data such as for example the current fuel level.

On top of this come specially developed applications with which the driver or rider can trace and share their routes with their community or download routes shared with the community. The make navigation functions and the latest traffic information available. The pool of apps is being continuously extended – nationally and internationally. Developers not only customize already existing apps for mySPIN, but also widen the app pool with new innovative ideas.

Manufacturers have, thanks to the cloud connectivity of mySPIN, the opportunity to address the driver or rider directly and keep them up to date with important, personalized information. Manufacturers themselves decide on the availability of the apps and their scope and can extend the portfolio to include new functions and apps.

Benefits for the user

– Comfortable use of apps while driving or riding without the driver or rider being distracted.
– Easy operation via touchscreen in the vehicle or via the handlebar control unit on the motorcycle.
– A large selection of apps from different categories.
– Compatible with iOS and Android operating systems.
– Driver- and rider-friendly display for optimal operation.
– Extensive customization options for vehicle manufacturers.
– Flexibility with regard to choice and scope of apps and additional functions in the vehicles.
– Access to vehicle data for data analyses – possibility of integration of cloud services.
– Real-time information on dangers and state of roads.
– Offline use of the navigation apps.
– Wide range of smartphone apps.
– Selected and optimized apps specially for the two-wheeler and powersports segment.
– Intuitive and safe use of smartphones and apps in the motorcycle and powersports segment: for more safety, comfort, and fun on the road.

Display and control

Interaction channels

Drivers have to process a constantly increasing stream of information originating from their own and other vehicles, from the road, and from telecommunication equipment. All this information must be conveyed to drivers on suitable display and indicating equipment that complies with ergonomic requirements.

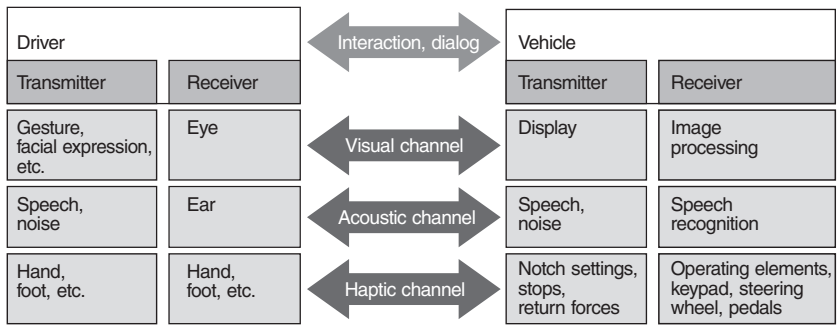
Visual channel – seeing
A human being discerns their environment predominantly through their sense of sight (Figure 1). Other road users, their position, their suspected behavior, the lane, and objects in the road area are detected with the human being's sight apparatus and their subsequent highly powerful image processing and interpretation, selected and evaluated by further structures in the brain with regard to their development and relevance.

Road-traffic infrastructure makes demands on above all the visual channel: Road signs communicate rules, markings delimit lanes from each other, direction-indicator lamps indicate a change of driving direction, stop lamps warn against vehicles that are slowing down. The visual channel is therefore extremely important to the driving environment. This applies to conscious vision, whereby the driver turns their eyes specifically to objects and focuses on them, but also to peripheral vision, which is crucial to positioning the vehicle within the lane. For this reason, additional glances at displays in the vehicle during interaction with driver-information and driver-assistance systems or their monitoring must be carefully assessed as to possible effects on road safety.

Acoustic channel – speaking and hearing
For communication with other road users, particularly the display and signaling of danger, the acoustic channel is used by human beings and by driver-assistance systems. In the driver's own vehicle this channel is also used to input commands via speech input systems and to output warnings and information from the driver-assistance system to the driver with acoustic signals and speech output.

The input of voice commands does not require the driver to turn their eyes, but does take up the driver's attention. Hearing does not require the driver to turn their eyes, but can convey spatial and complex information (e.g., description of the situation at intersections) poorly. Even drivers who do not have an adequate sense of hearing must be able to communicate with the driver-assistance system.

Figure 1: Information channels between driver and vehicle



Haptic and kinesthetic channels – operation and feeling

The haptic channel provides feedback to the driver in the case of all motor operations when switches are operated and during steering and braking. Warning the driver by tightening their seat belt briefly and by vibrating the seat in the event of an imminent departure from the present lane has already been introduced in series production. A further possibility for arousing the driver's attention is the vibration of the steering wheel. Increased resistance on the accelerator pedal can support a speed recommendation and heightens the driver's attention. A torque that the driver can feel can be exerted on the steering wheel to suggest an evasive maneuver.

Even the kinesthetic channel, which helps the driver to perceive instances of acceleration while driving, is already in use in production vehicles for the purpose of arousing the driver's attention, for instance by way of a short braking jolt.

The haptic channel is however continuously required for steering and cruise control, and additional manual operations (e.g., operation of the cellular phone's keypad) may impair lateral guidance by distracting the driver.

Instrumentation

Information and communication areas

In any vehicle, there are four information and communication areas which must satisfy different requirements in terms of their display features: the instrument cluster, the windshield, the center console, and the rear cabin.

The supply of available information and the necessary, advisable or desirable information for the respective occupant determine which communication areas are used in the case. Dynamic information (e.g., driving speed) and monitoring information (e.g., fuel level), to which the driver should respond, is displayed in the instrument cluster in a favorable reading position close to the driver's primary field of vision.

A head-up display (HUD), which reflects the information on the windshield, is ideally suited to engaging the driver's attention (e.g., in the case of warnings from a driver-assistance system or for route directions). The head-up display is also suitable for displaying the speed in a digital form. A supplementary feature is the output of acoustic signals or speech.

Status information or more extensive operation dialogs (e.g., for vehicle navigation) are preferably shown in the central display in the center console. The operating units are usually assigned to this central display for more compact vehicles. For more spacious vehicles with a high-mounted central display (see Figures 2c and 2d), the preferred place to arrange the control units is in the primary reach area in the form of a rotary/push control element on the transmission hump.

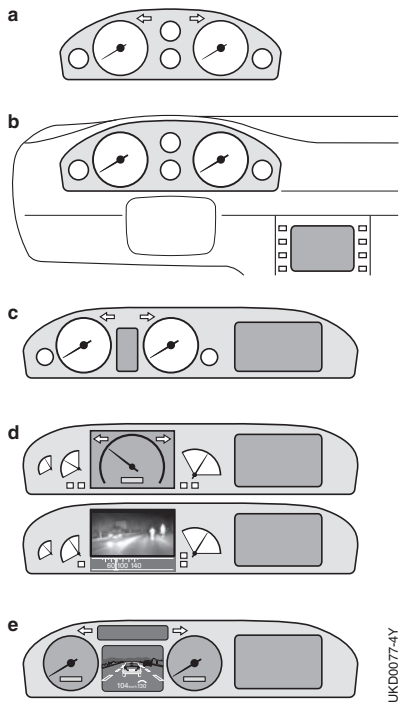
Information of an entertainment nature is featured in the vehicle rear compartment, away from the primary field of vision. This is also the ideal location for a mobile office. The backrest of the front-passenger seat is a suitable installation location for the monitor and operator unit of a laptop computer.

"Dual View" or "Split View" displays for the center console have additionally been available since 2009; these displays can show different information to driver and passenger so that the passenger can, for example, watch videos without distracting the driver while the driver sees the navigation-system information.

Instrument clusters
Earlier single instruments for the optical output of information (e.g., driving speed, engine rpm, fuel level, and engine temperature) were initially superseded by more cost-effective, well-lit and antireflective instrument clusters (combination of several information units in one housing).

Figure 2: Driver information area with instrument cluster and monitor in the center-console area (development steps)
a) Needle instruments,
b) Needle instruments with TN LCD and separate AMLCD in center console,
c) Needle instruments with integrated (D)STN LCD or AMLCD and AMLCD-technology monitor in center-console area,
d) Freely programmable instrument cluster with two AMLCD components with different display options,
e) Two freely programmed liquid-crystal displays.

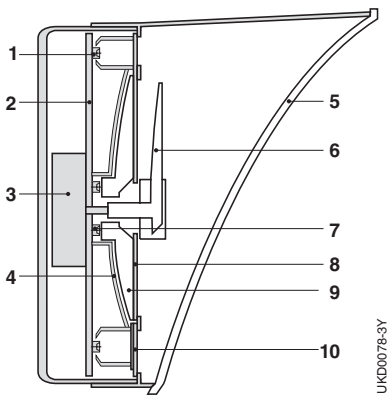
For designations, see Display types.



The passage of time, with the continual increase in information, saw the creation of the modern instrument cluster in the existing space available, with several needle instruments and numerous indicator lamps (Figure 2a). This first generation of instrument cluster was dominated by the eddy-current measuring instrument for the speedometer, the moving-coil measuring instrument for the rev counter, and the hot-wire instrument for the fuel level.

Measuring instruments
Even today most instruments still operate with mechanical needles and dial faces (Figures 2a to 2c). Initially moving permanent-magnet ratiometer elements replaced the voluminous, mechanically driven eddy-current measurement instruments. Today stepping motors with low overall depth and high torque are predominant. Due to a compact magnetic circuit and (mostly) two-stage gearing with power consumption of only approx. 100 mW, these motors permit swift and highly accurate needle positioning (Figure 3).

Figure 3: Instrument cluster (design)
1 Indicator lamp (LED),
2 Printed-circuit board,
3 Stepping motor,
4 Reflector,
5 View cover,
6 Needle,
7 Backlighting with backlit-technology LED,
8 Dial face,
9 Optical waveguide,
10 LCD (odometer).



Digital displays
The digital instruments fitted up to the 1990s (e.g., for indicating driving speed) displayed information using vacuum-fluorescence display (VFD) technology and, later, liquid-crystal (LCD) technology, but they have now largely disappeared in Europe. This unfortunately means that the advantage of a demonstrably fast and accurate speed reading has been lost. Today's Formula 1 cars all have digital speedometers.

Digital displays in instrument clusters can be found today in Japan (e.g., Toyota Prius) and in luxury-class cars in the USA.

Lighting
To illuminate instrument clusters, backlighting technology has, on account of its attractive appearance, gained acceptance to a large extent over the frontlighting technology used previously. Bulbs have been replaced by long-lasting light-emitting diodes (LEDs). LEDs are suitable both as warning lamps and as backlighting for scales, displays and needles (using plastic optical waveguides if necessary).

Where for a long time only monochrome LEDs in the colors yellow, orange and red were available, now blue and highly efficient white LEDs are available.

Color LCDs, on account of their very low transmission (typically approx. 6 %), had initially required cold cathode fluorescent lamps (CCFLs) for the backlighting in order to achieve a good contrast in daylight. These have in the meantime been replaced by highly efficient white LEDs.

Graphics modules in the instrument area
Fitting vehicles with a driver's airbag and power-assisted steering as standard has resulted in a reduction in the view through the top half of the steering wheel. At the same time the amount of information that has to be displayed in the installation space available has increased. Prior to the availability of graphics displays whose display surfaces can show any information – flexibly and ordered by priority – engineers tried to solve this installation dilemma initially using additional displays which were positioned on the dashboard – e.g., trip computers and the first monitors for displaying navigation information, which were sometimes also mounted on a gooseneck. The availability of graphics monitors resulted in the hybrid instrument with classic needle instruments and a graphics module (Figure 2c).

The graphics module in the instrument cluster mainly enables an indication of driver-related functions such as, for example, service intervals, check functions covering the vehicle's operating state, as well as vehicle diagnostics as needed by a workshop. They can naturally also show route-direction information from the navigation system, where, due to the small surface area of these modules, digitized map excerpts must be missed out so that only route-direction symbols such as turnoff instructions or intersection symbols appear in the display. This limitation has also been retained in newer vehicles. The initially monochrome modules were quickly followed in higher-specification vehicles by color displays which can be read more quickly and accurately because of their color resolution.

Since 2005 active-matrix displays (see below) have also been in use to represent analog instruments (Figure 2d). One example is the instrumentation in the Mercedes S-Class, in which the speed indication is emulated using an LCD monitor. During changeover to the night-vision function (see Night-vision system) the camera's conditioned video image appears, while the engine rpm indication is shown in bar form under the video image.

With the launch of the Mercedes S-Class in model year 2013 the instrument cluster was replaced for the first time by a full-screen liquid-crystal monitor. Such a freely programmable instrument offers many benefits. Either the information can be arranged according to customer preference (e.g., at the end of the line) in the display or drivers who are familiar with digital media can program their instrument clusters themselves. This option is offered in the Chevrolet Corvette C7.

In view of the drop in prices of these displays it can be assumed that the digital display in the instrument cluster will become widely accepted because in this way the variety of versions of entire vehicle model series can be reduced to a single, flexibly programmable unit, resulting in huge logistical savings.

The introduction of the graphics monitor as the instrument cluster was also accompanied by a growth in the format of the center-console display.

Central display and operating unit in the center-console area

The arrival of driver-information systems, initially mainly navigation and telephone, saw the establishment of a further monitor with accompanying operating elements in the center console (Figure 2). It was initially positioned here because it was possible here to create sufficient installation space with the installation bay for the car radio and the ashtray. The operating elements were arranged along the edges of the display (Figure 2b). The vehicle manufacturers have since switched to the ergonomically more favorable solution of installing the monitor in the center-console area at the height of the instrument cluster (Figure 2c). In this connection, most German automobile manufacturers have replaced the button input on the monitor edge with a combined rotary/push input button in the driver's primary area of reach on the center console. Operation directly via the display is possible with touch screens.

A wealth of information is shown on this central monitor, e.g., navigation, communication (telephone, SMS, Internet), audio (radio, digital sound recording media), video and TV when the vehicle is stationary, climate control, and settings (date, time). Information from the new driver-assistance systems (e.g., parking and maneuvering aids) can also be usefully displayed here.

The important issue for all visual displays is that they can be easily read within the driver's primary field of vision or its immediate vicinity so as to ensure that the driver does not have to look away from the road for long periods. Positioning this terminal, which is of universal use to driver and passenger, in the upper center-console area is effective and necessary from both ergonomic and technical standpoints. The demands placed by TV reproduction and the navigation system on the video and map display determine the demands placed on the resolution and the color reproduction of the monitor.

For this central display monitor, there has initially been a change from a small 6" monitor with a 4:3 aspect ratio to a wider format with a 16:9 aspect ratio, which allows additional route-direction symbols to be displayed as well as the map. Newer vehicles with the central monitor located in the upper center-console area sometimes have, due to the greater space available, much larger monitor formats. The Tesla Model S even uses a vertically arranged monitor with a diagonal of 17" as a combined display and operating unit in the center console.

Display types

Liquid-crystal display
The most display most commonly used today in vehicles is the liquid-crystal display (LCD). Being a passive display which emits no light itself but modulates incident light, it requires additional lighting.

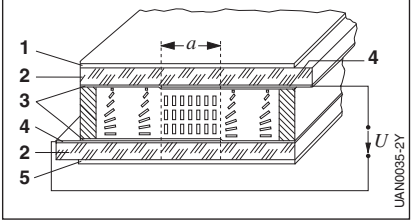
Twisted Nematic LCD
TN LCD (Twisted Nematic Liquid-Crystal Display) is the most commonly used form of display. The name derives from the arrangement internally of rectangular molecules twisted through 90° at zero voltage. This causes the polarization plane of the light passing through the cell to rotate through 90° (Figure 4). If this cell is arranged between crossed polarizers, it is transparent in the de-energized (zero-voltage) state. When a voltage is applied, the molecules are rotated in the field direction (Figure 4, middle segment). The cell becomes non-transparent, i.e., it is a light valve.

The operating range is –40 to 125 °C. The switching times are slightly longer at low temperatures due to the viscous properties of the liquid-crystal material.

Depending on the choice of polarization direction of the externally applied polarizer film, the TN LCD can be operated in positive contrast (dark characters on a light background) or negative contrast (light characters on a dark background). The TN LCD is usually operated in negative contrast in the vehicle since this comes close to the familiar appearance of printed dials. Negative-contrast cells always required backlighting.

Separately controlled segment areas can be used to portray numbers, letters, and symbols. TN technology is suitable not only for smaller display modules but also for larger display areas and modular LCD instrument clusters such as the instrument cluster in the Audi Quattro model year 1984.

Figure 4: Principle of operation of a liquid-crystal display (nematic cell)
1 Polarizer, 2 Glass,
3 Orientation and insulation layer,
4 Electrode, 5 Polarizer.
a Segment area.



Active-matrix LCD
Dot matrix displays with graphics capabilities are needed to display infinitely variable information. The task of the visually sophisticated and rapidly changing display of complex information in the area of the instrument cluster and the center console with high-resolution liquid-crystal monitors with video capabilities can only be performed effectively by an active-matrix liquid LCD (AMLCD). The pixels are addressed with thin-film field-effect film transistors (TFT LCD, Thin-Film Transistor LCD). Display monitors with diagonals of 3.5 to 10 inches in the center-console area and an extended temperature range (-25 to 110 °C) are available for motor vehicles and are the most widely used monitor technology in vehicles today. Formats of 10 inches and more are used for the freely programmable instrument cluster.

AMLCDs (Figure 5) consist of the "active" glass substrate with the semiconductor structures and the opposing plate with the color-filter structures. The active substrate accommodates the pixel electrodes made from indium tin oxide, the metallic row and column conductors, and semiconductor structures. At each intersecting point of the row and column conductors, there is a thin-film field-effect transistor, which is etched in several masking steps from a previously applied sequence of layers. A capacitor is generated at each pixel as well.

The opposite glass plate accommodates the color filters and a "black matrix" structure for covering the metallic row and column conductors, which improves display contrast. These structures are applied to the glass in a sequence of photolithographic processes. A continuous counter-electrode is applied on top of them for all the pixels. The color filters are

applied either in the form of continuous strips (good reproduction of graphics information) or as mosaic filters (especially suitable for video pictures).

Organic light-emitting diodes

The luminescence excitation for organic, electroluminescent display units (OLED, organic light-emitting diode) occurs, as in light-emitting diodes (LEDs), through injected charge carriers, but here in a layer of organic materials. Three processes have to take place within the system of coatings for electroluminescence to occur: Injection, transportation of both types of charge carriers (electrons, holes) through the layer as well as their recombination.

The organic layers are embedded between two electrodes. Figure 6 shows the basic structure of an OLED picture element. Glass (or plastic foil) coated with indium tin oxide (ITO) is typically used as the anode; this lets the generated light through (downwards in the graphic). The cathode is made of metal. The organic material is located between them. Charge carriers are injected into the material using the electric contact points by applying a voltage. These are moved toward each other by the polymer layer. When charge carriers of different charges meet (electrons and holes), this causes the radiative recombination as described above (see also OLED, electronics).

Figure 7 shows the structure of an OLED matrix display. The organic layer is located between the row and column electrodes arranged perpendicularly to each other. Multi-color reproduction is achieved by OLED materials in a dot matrix that emit different colors.

OLED displays have already been established in the consumer area (e.g., mobile phones) for years. Due to the high requirements for resistance to climatic conditions, OLEDs have only been used for small, monochrome, alphanumeric and simple graphic displays to date, e.g., for

Figure 5: Opened AMLCD

1 Color filter (blue, red, green),
2 Counter-electrode, 3 Glass substrate,
4 Column conductor, 5 Pixel electrodes,
6 Row conductor,
7 Thin-film field-effect transistor,
8 Black matrix structure.

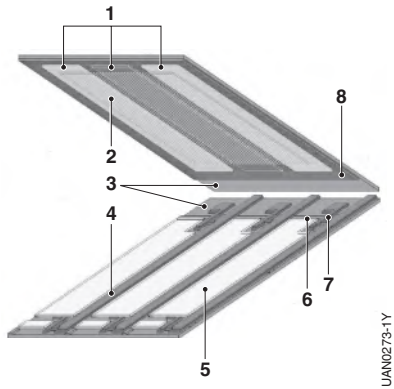


Figure 6: Structure of an OLED picture element

1 Metal cathode (Al, In),
2 Polymer or monomer layer system,
3 Transparent anode,
4 Substrate made of glass or flexible plastic foil (PET).

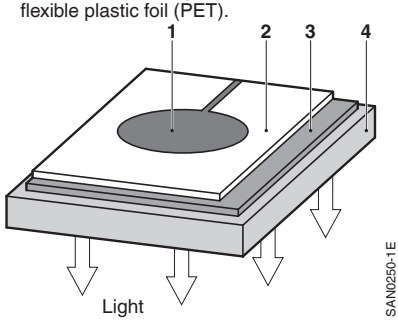
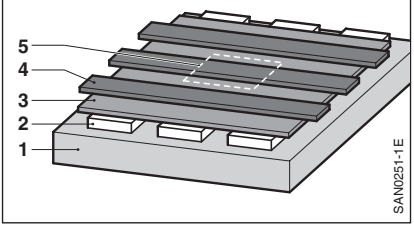


Figure 7: OLED matrix display

1 Substrate made of glass or flexible plastic foil (PET),
2 Anode strip,
3 Organic layer system,
4 Cathode strip,
5 Pixels.



displaying air-conditioner functions. The option to construct OLEDs with curved surfaces using plastic substrates could attract interest in the future from the standpoint of providing good anti-reflection conditions in a vehicle.

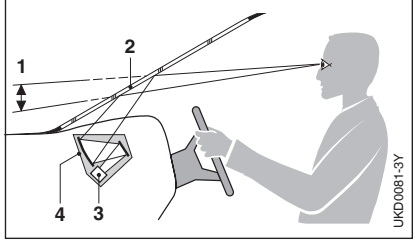
Head-up display

Conventional instrument clusters are situated at a viewing distance of 0.8 to 1.2 m. In order to read information in the area of the instrument cluster, the driver must adjust his or her eyes from long distance (observing the road scene) to the short viewing distance for the instrument. This accustomization process usually requires, depending on the driver's age, 0.3 to 0.5 s.

Head-up displays (HUDs) were first used in military-aviation applications. For motor-vehicle applications they were available in a simple design, usually as a digital speedometer, in parallel with the analog speedometer in Japan and in the USA as an optional extra. An improved form was initially used in the USA and later also in European vehicles.

The image of a head-up display is projected with the aid of the windshield into the driver's primary field of vision. The optical system generates a virtual image at such a viewing distance that the human eye can remain adjusted to long distance. Head-up displays do not require the driver to divert their eyes from the road – thus allowing the driver to register at all times safety-critical driving situations without being distracted. The driver does not need to look at the instrument cluster when important information is shown in the head-up display.

Figure 8: Head-up display (principle)
1 Virtual image,
2 Reflection in the windshield,
3 Display module: liquid-crystal display (LCD) with backlighting,
4 Optical system.



Design

A typical head-up display (Figure 8) features an activated display module for generating the image, a lighting facility, an optical imaging unit, and a combiner, which reflects the image to the driver's eyes. In a motor vehicle it is usually the windshield which acts as the combiner. To avoid double images – caused by reflections on the inner and outer boundary surfaces – the windshield is slightly wedge-shaped in design. From the driver's perspective the two images created at the boundary surfaces then coincide.

A real image is created in the display module. This module can be a display or a back-projection onto a diffusing surface, e.g., with a scanning laser beam. This real image is only roughly 20×40 mm in size, but the luminance required for good readability in daylight is more than 50,000 cd/m^2, i.e., 100 times more than a conventional direct-view display. This image is reflected through the windshield into the driver's eyes. For the driver this virtual image of the driving situation in front of the vehicle is superimposed.

Optical elements (lenses, concentrating reflectors) are generally inserted into the beam path to increase the virtual distance of the image.

Particularly high-contrast segment TN LCDs can be used as the display for monochrome head-up displays with modest levels of information content. More sophisticated, multicolor displays use polysilicon-technology AMLCDs.

Display of information of head-up displays

The virtual image should not cover and obscure the road scene ahead; it is therefore displayed in an area with low information content, in other words "floating above the hood" (Figure 9). In order to prevent the driver from being overwhelmed with stimuli in his or her primary field of vision, the head-up display should not be overloaded with information, and is therefore never a practical substitute for the conventional instrument cluster. It is, however, particularly well suited for

displaying navigation notices (arrows) and safety-related information such as warnings or information from the ACC system (Adaptive Cruise Control, see Adaptive cruise control system).

Because of the beneficial display of information, it can be assumed that the number of head-up displays worldwide will increase dramatically.

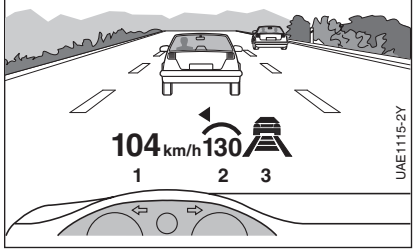
Augmented-reality head-up display

"Augmented-reality" head-up displays offer an even greater benefit than showing information in the head-up display without local reference to the environment. They are called "augmented-reality displays" in the English-speaking world. They require a greater level of technical sophistication in realizing the projection unit by using a laser or a DLP module to generate the image. DLP stands for "Digital Light Projection", the generation of images by micromechanical elements, as are also used in projectors.

Augmented-reality head-up displays are characterized by a larger projection field. This enables the system to display the warning of an obstacle in the driver's field of vision and at the virtual distance in which it is located in nature. To enable the information to be read without parallax, however, a position correction depending on the position of the driver's head is required. With this system, route-direction information can be displayed for example at an intersection precisely where the junction suggested by the navigation system is situated; at an expressway/highway exit the arrow for recommending the exit can be positioned on the deceleration lane [1].

Figure 9: Information shown in the head-up display

1 Present speed,
2 Desired speed,
3 ACC status (Adaptive Cruise Control).



Touchscreen

The infotainment-system display is a touch-sensitive screen (touchscreen). Most of the functions of for example the radio, the navigation system or the smartphone paired by Bluetooth with the infotainment system can be operated by tapping the display with a finger.

Voice control

Voice input can be used to enter addresses in the navigation system or enter contacts that the user wishes to call on the smartphone paired with the infotainment system. Modern in-car voice-control systems have taken a tremendous technological step forward in recent years thanks to the integration of artificial intelligence, “natural language understanding” (NLU), and “cognitive computing”. Communication between driver and system is becoming more natural and speech-recognition capabilities have improved, even when it comes to understanding regional accents.

Conventional speech-recognition systems, like the ones in smartphones, always require a fast data connection to process voice commands. Speech is recorded and then transmitted to an external computer system for analysis. This method is only of limited use for in-car applications. After all, a fast data connection is not always available out on the road. Speech recognition must for this reason function without dependence on an external data connection. The on-board infotainment system handles all the computing operations. The voice thus remains a reliable means of control, even in tunnels or in well-developed areas with adequate cellular coverage.

Voice control for the infotainment system provides the following product benefits:
– The system is available everywhere because the software functions offline,
– recognition of regional accents thanks to innovative software,
– cultural and regional differences are taken into consideration,
– safe and comfortable driving owing to fewer distractions.

References
[1] M. Wheeler: HUD Systems: Augmented Reality is coming to your windshield. Photonics Spectra 02/2016.

Radio and TV reception in motor vehicles

Wireless signal transmission

Wireless transmission technology permits the simultaneous broadcasting of information to large populations. It is also of major significance to mobile radio reception as used in motor vehicles. At present, the importance of digital transmission methods is growing. The wireless section of the transmission chain also uses analog signal transmission, so the two technologies are basically identical.

Radio and TV broadcasting

Wireless radio and TV broadcasting is primarily used for terrestrial transmissions. In the case of analog radio broadcasting, the audio signal is modulated on the high-frequency signal. In the receiver, the received high-frequency signal is then converted to the baseband frequency and demodulated. The final signal obtained in this way is identical to the useful signal.

In telecommunications, the propagation of an electromagnetic wave is used to transmit information whereby the amplitude, the phase or the frequency of this oscillation is altered depending on the information to be transmitted. The customary frequencies range from a few kilohertz up to 100 GHz. Some of the commonly used frequency ranges are set out in Table 1. The use of frequency bands is subject to legal control (Telecommunications Act for Germany, 2004). Each country's frequency allocation plan is based on international agreements that are set out in Article S5 of the Radio Regulations of the ITU (International Telecommunication Union, [1]).

Information transmission using highfrequency waves

The variation in a high-frequency signal used to transmit a useful signal from a transmitter to a receiver is called modulation. The modulated high-frequency signal is emitted within a precisely determined, narrowly defined frequency band by an antenna. The receiver selects precisely that frequency band from the large number of frequencies received by the antenna. In this way, wave propagation between transmitter and receiver is a link within the signal transmission chain.

For example, when it comes to the transmission of an audio signal, the useful signal in contrast to the high-frequency carrier signal is made up of various frequencies ranging up to a maximum of 20 kHz. The high-frequency carrier is modulated with this low-frequency signal. A transmitter antenna beams out the carrier wave.

Table 1: Overview of some frequency bands in radio and TV broadcasting

Wavebands	Wavelengths λ in m	Frequencies f in MHz	Examples
Long wave (LW)	≈ 2,000 – ≈ 1,000	0.148 – 0.283	Analog radio broadcasting
Medium wave (MW)	≈ 1,000 – ≈ 100	0.526 – 1.606	Digital Radio Mondiale (DRM)
Short wave (SW)	≈ 100 – ≈ 10	3.950 – 26.10	
Very high frequency (VHF)	≈ 10 – ≈ 1	30 – 300	
Band 1		47 – 68	TV
Band 2		87.5 – 108	Digital Audio Broadcasting (DAB)
Band 3		174 – 223	
Ultra high frequency (UHF)	≈ 1 – ≈ 0.1	300 – 3,000	
Band 4		470 – 582	TV
Band 5		610 – 790	Digital Video Broadcasting (DVB-T2), Digital Video Broadcasting (DVB-H)
L-band		1,453 – 1,491	Digital Audio Broadcasting (DAB)
Super high frequency (SHF)	≈ 0.1 – ≈ 0.01	3,000 – 30,000	
		10,700 – 12,750	Digital Video Broadcasting (DVB-S)

The maximum distance at which the signal can still be received and the reception quality depend on the frequency, among other things. Short-wave and long-wave have very long ranges, in some cases intercontinental, whereas the reception range of VHF transmissions is hardly any further than the line of sight.

The receiver station demodulates the signal. The resulting low-frequency electrical oscillation is then converted by a loudspeaker into acoustic oscillations.

Amplitude modulation
Amplitude modulation (AM) is the change in amplitude A_H of the high-frequency oscillation at frequency f_H in synchronization with the low-frequency oscillation (A_N, f_N) (Figure 1).

Amplitude modulation is used in the short-wave, medium-wave and long-wave bands, for example.

Frequency modulation
Frequency modulation (FM) is the change in frequency f_H of the high-frequency oscillation in synchronization with the low-frequency oscillation (Figure 2).

Frequency modulation is used for FM radio and the sound channel of analog TV transmissions, for example. The transmissions of frequency-modulated signals are impaired by amplitude-modulation interference (e.g. caused by the ignition systems of spark-ignition engines) to a lesser degree than the transmissions of amplitude-modulation transmitters.

Digital modulation processes
In the case of digital modulation processes, the amplitude or the frequency of the carrier is discretely altered. One or more bits can thus be allocated to each of these carrier states so as to enable digital information to be transmitted.

Reception problems
VHF signals have virtually a linear propagation. As a result, a car radio may lose the signal from a VHF transmitter that is only 30 km away if there is high ground between the car and the transmitter. On the other hand, reception may be unimpaired at a location further away if there is a clear "line of sight" between the car and the transmitter. Such "radio shadows" are thus often covered by a fill-in transmitter.

Signals may be reflected off the sides of valleys or high-rise buildings. The reflected signals then arrive at the receiving antenna with a time delay and are superimposed on the signals received directly from the transmitter. This produces what is known as "multi-channel reception". It causes multipath interference and results in the deterioration of sound quality in radio reception.

Propagation of electromagnetic waves is impaired by conductors within the transmitter's radiation field (e.g. steel masts or power lines) and even nearby forests, houses, or locations in deep valleys. The characteristics of wave propagation are important when it comes to suppressing interference in motor vehi-

Figure 1: Amplitude modulation (AM)
a) Low-frequency oscillation at amplitude A_N and frequency f_N,
b) Unmodulated high-frequency oscillation,
c) Modulated high-frequency oscillation.

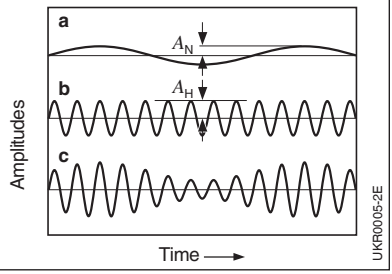
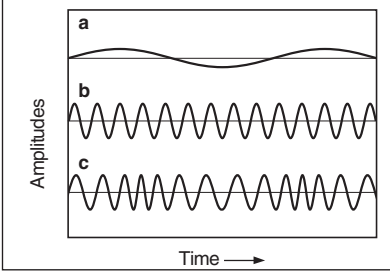


Figure 2: Frequency modulation (FM)
a) Low-frequency oscillation at amplitude A_N and frequency f_N,
b) Unmodulated high-frequency oscillation,
c) Modulated high-frequency oscillation.



cles effectively. Interference-free reception is impossible if the signals received from the transmitter are too weak. Thus, reception of a previously perfectly interference-free signal will break down suddenly when a car enters a tunnel, for example. This can be explained by the shielding effect of the reinforced-concrete tunnel walls which reduce the useful field strength of the transmitter signal to which the radio is tuned. At the same time, the interference-field strength remains the same. Under certain conditions, it may not be possible at all to continue to receive the radio-station signal. Similar phenomena can also be experienced when traveling in mountains, for example.

Radio interference

Radio interference is caused by undesirable high-frequency waves which are directed together with the desired signal to the receiver. It occurs wherever electrical currents are suddenly interrupted or switched on. High-frequency interference waves are thus produced, for example, by the ignition of a spark-ignition engine, the actuation of a switch, or the switching operations on the commutator of an electric motor. Such rapid changes of current generate high-frequency waves which interfere with radio reception by receivers located close by. The effect of the interference depends on the steepness of the signal pulse and its amplitude, among other things.

Radio interference caused by such steeply rising current pulses can be reduced or completely eliminated by EMC measures (electromagnetic compatibility).

Interference can travel to the receiver in different ways: directly through wires connecting the interference source and the receiver, or by wireless transmission of electromagnetic radiation, or by capacitive or inductive coupling. Strictly speaking, the last three options cannot be separated from one another.

Signal-to-noise ratio

Reception quality depends on the strength of the electromagnetic field generated by the transmitter. This should be substantially greater than the strength of the interference field, i.e. the ratio between the strength of the transmitter signal and the strength of the interference field – the signal-to-noise ratio – should be as large as possible.

A receiver close to the interference source receives not only the useful signal from the desired transmitter, but also the undesired interference signal, if it is transmitted at the same frequency. Nevertheless, good-quality reception is still possible provided that the field strength of the desired transmitter signal at the point of reception is very high in comparison with the strength of the electromagnetic field generated by the interference source. The useful field strength of the signal from the transmitter depends on transmitter power, transmitter frequency, the distance between the transmitter and the receiver, and the propagation characteristics of the electromagnetic waves. In the case of medium-wave and long-wave signals, the field strength of the transmitter signal can be weakened by difficult topology to such an extent that the signals of even powerful transmitters have low useful field strengths at some reception locations. VHF signals may be subject to heavy fluctuations in useful field strength under certain conditions. Receivers in motor vehicles may also suffer from relatively low useful voltages at the receiver input due to the short effective height of the antenna. Accordingly, the opportunities for improving the signal-to-noise ratio in the receiver are very limited.

By optimizing antenna positioning, the available useful voltage at the receiver input can be increased, thereby improving the signal-to-noise ratio, which is the decisive factor in reception quality. Frequently, however, a compromise is made between design considerations and technical demands. Another means of improving signal-to-noise ratio is by reducing the strength of the radiated interference signals.

Receiver design also has an impact on reception quality. In addition to metallic shielding, which prevents the direct entry of radiated interference, and filters on the power source input, some receivers have circuits fitted with automatic interference suppression (see section entitled “Reception improvement”).

Radio tuners

Radio tuners in motor vehicles are frequently called car radios or automotive sound systems. However, this term refers not only to a radio tuner but also to devices with a large number of integrated information and entertainment functions. These include, for instance, the analysis of supplementary information (e.g. traffic news), players for reproducing from storage media (e.g. CDs and SD cards), as well as integrated radio communication interfaces, cellular phones and other devices.

In the past few years, conventional analog transmission technology has developed new systems. For this reason, modern car tuners are capable of receiving a wide variety of radio broadcast systems worldwide. As well as conventional radio broadcasting, these systems include among others DAB (Digital Audio Broadcasting), DRM (Digital Radio Mondiale) and SDARS (Satellite Digital Audio Radio Services).

The conventional radio tuner is designed to receive analog FM and AM modulation and has an analog signal path from the antenna to the audio signal. Modern car tuners with the highest received power, on the other hand, process signals digitally. The IF signal (intermediate frequency) supplied by the tuner is digitized by an analog-digital converter, and then processed. In standards covering digital modulation essentially only demodulation in the digital part of the signal path changes.

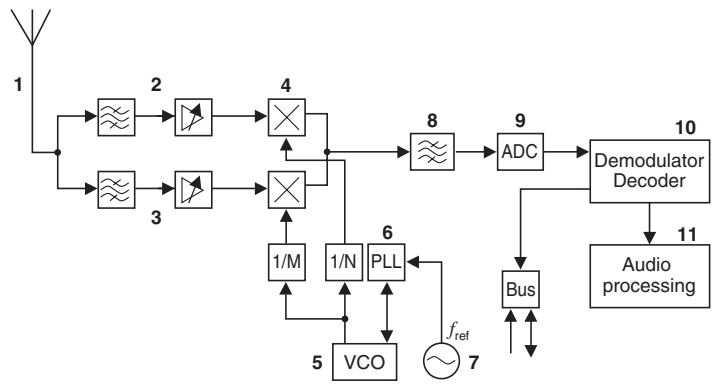
Conventional tuners

Signal processing

The antenna, which is predominantly designed as a rod or window antenna, picks up the electromagnetic signal emitted by the transmitter. This consists of different channels with a fixed frequency spacing. The high-frequency alternating voltage generated at the base of the antenna is routed to the receiver and processed there.

The conventional tuner for analog radio reception basically contains two signal paths: one for processing amplitude-modulated signals and one for processing frequency-modulated signals. These are

Figure 3: Block diagram of a radio tuner
1 Antenna, 2 FM input stage, 3 AM input stage,
4 Mixer stage, 5 Voltage-controlled oscillator (VCO),
6 Phase-locked loop (PLL), 7 Reference-frequency oscillator,
8 Intermediate-frequency filter (IF), 9 Analog-digital converter,
10 Demodulator and decoder, 11 Audio processing.
f_{ref} Reference frequency.



usually divided into the blocks described in the following (Figure 3).

AM input stage
A bandpass filter limits the amplitude-modulated signal in the LW, MW and SW bands, and the resulting signal is amplified in the following stage with low noise.

FM input stage
The frequency-modulated VHF signal is received by a separate input stage. The input filter is either tuned to the frequency to be received or configured to the entire received band. The received-signal level is then adjusted to the desired input level for the following mixer stage with low noise by an automatically controlled amplifier.

Voltage-controlled oscillator
The oscillator (VCO, voltage-controlled oscillator), whose frequency is controlled by a phase-locked loop (PLL), generates a high-frequency oscillation which is subdivided. The input signal is converted with the aid of this signal into a constant intermediate frequency (IF) in the mixer stage. A quartz-stabilized signal serves as reference frequency.

Mixer stage
The mixer stage converts the input signal into a constant intermediate frequency. Different mixer stages are frequently used for receiving FM and AM signals. On the other hand, the principle of frequency conversion is the same.

IF filter and amplifier
The IF signal recovered in this way is then fed to an IF filter and a controlled amplifier.

Analog-digital converter (ADC)
The analog-digital converter (ADC) converts the analog IF signal into a digital signal.

Demodulator
The demodulator generates the digital audio signal from the digital IF signal.

Decoder
Additional information, such as RDS data (Radio Data System), is decoded by the decoder and passed on to a processor for processing.

Audio processing
After demodulation, the audio signal may be adapted, for example, to the vehicle conditions and the listener's preferences. This can be achieved by using the appropriate controls to adjust tone and volume or to vary the level between front and rear or between left and right.

Digital receivers
A digital receiver (ADR, advanced digital receiver) is a highly integrated receiver module whose input signal is the analog or digital IF signal. The analog signal is converted into a digital signal, the digital signal is processed on the digital level. This technology allows a type of signal processing that would not be possible using analog technology. It generates, for example, IF filters which provide for exceptionally good harmonic-distortion levels, and whose bandwidth is variable and can be adapted to the reception conditions. In addition, there is a larger number of other means of processing the received signal in such a way as to substantially reduce interference in the audio signal (see SHARX, DDA and DDS).

Digital equalizer
The digital equalizer (DEQ) consists of a multiband parametric equalizer which provides separate adjustment features for the mid-frequency, and the amplification/attenuation of individual filters. The suppression of unwanted resonances thus obtained optimizes sound quality inside the vehicle. The frequency response of the speakers can also be linearized.

Preset equalizer filters can also be implemented in some devices. These can be activated according to music genre or vehicle type (e.g. jazz or pop, van or sedan).

Digital Sound Adjustment
"Digital Sound Adjustment" (DSA) is a system which automatically analyzes and corrects frequency response in in-car systems. A microphone and a digital signal processor (DSP) are used to pick up and analyze a test signal generated by the speakers. The optimum sound curve for the vehicle is then set on the equalizer.

Dynamic Noise Covering
While the vehicle is in motion, the "Dynamic Noise Covering" (DNC) function uses a microphone to constantly detect and analyze the vehicle-noise spectrum that is blocking out the audio signal and impairing the perceived sound quality. The function uses selective amplification or selective dynamic compression (reduction of dynamic response between minimum and maximum values) of the interference frequencies to maintain optimum reproduction of sound quality, regardless of driving noise.

Reception quality
Analog radio broadcasting is primarily used for terrestrial transmissions. The transmission path is not always ideal, with the result that reception quality may be impaired, depending on the transmitter/receiver constellation and environmental conditions. In the case of VHF reception, reception locations may be critical because of the transmission-path problems described below.

Fading
Fading is caused by fluctuations in signal reception level due to obstacles in the signal path, such as tunnels, high-rise buildings, or mountains.

Multipath reception
Multipath reception is caused by the reflection of signals off buildings, trees or water. This can easily lead to a substantial dip in reception field strength, and even total signal loss. The differences in the strength of the received signal field strength occur within a few centimeters of each other. Such fluctuation has a particularly detrimental effect on mobile receivers, such as car tuners.

Adjacent-channel interference
Adjacent-channel interference occurs when there is another channel received with a high field strength close to the channel received.

High-level signal interference
High-level signal interference occurs close to the transmitter at high field strength. The receiver protects its input by reducing the field strength. This has the effect of attenuating the weaker signals from other transmitters, i.e. they become quieter.

Overmodulation
Some transmitters increase the modulation level in order to achieve a greater range or higher loudness level. The disadvantages of this process are greater distortion factor and greater susceptibility to multipath interference.

Ignition interference
High-frequency interference sources (such as the ignition of a spark-ignition engine, the operation of a switch, or switching operations on the commutators of electric motors) produce reception interference.

Reception improvement
A large number of functions for improving received power are implemented in modern automotive sound systems. An overview of the most important functions is provided in the following.

Radio Data System
Radio Data System (RDS) is a digital data transmission system for FM radio. This format is standardized throughout Europe. It provides the tuner with extra information in addition to the desired audio signal by using alternative received frequencies with the same modulation. This enables, for example, the tuner to switch continuously to the frequency with the least interference. Table 2 provides an overview of the information transmitted.

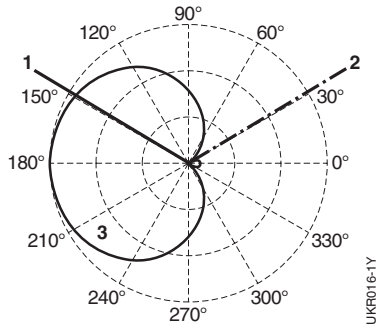
Table 2: RDS codes

Code	Information transmitted
PS	Name of the station received
AF	List of alternative frequencies on which the station also broadcasts
PI	Identifies the broadcast station
TP, TA	Traffic-program station identification, traffic-announcement identification
PTY	Identification of the station type
EON	Signals traffic announcements on a parallel station
TMC	Standardized traffic information
CT	Time (for synchronizing the clock in the vehicle)
RT	Text transmission (e.g. title of music track)

Digital Directional Antenna
The "Digital Directional Antenna" (DDA) system developed by Bosch uses the signals from two antennas to calculate a synthetic antenna with a new directional characteristic. This permits suppression of interference caused by multipath reception as shown in Figure 4.

Digital Diversity System
The reception characteristics of FM radio are heavily dependent on location. A Digital Diversity System (DDS) has a number of antennas at its disposal so that it can switch between them in order to enhance reception. The Digital Diversity System integrated in the digital receiver uses exactly the same signal for its switching strategy and the signal is available as the audio signal after demodulation.

Figure 4: Optimized antenna signal
1 Direct signal,
2 Reflection,
3 Directional diagram.



High-cut
Interference caused by such effects as fading and multipath reception has a greater effect on higher audio frequencies. For this reason, modern automotive sound systems have a facility for detecting such interference and reducing the level of the audio signal at higher audio frequencies when there is interference.

SHARX
SHARX is a function which automatically adjusts the bandwidth of the intermediate-frequency filter for FM reception to suit reception conditions. If different stations transmit at frequencies that are very close to one another, this function significantly increases the clarity of separation by reducing the bandwidth and allows virtually interference-free reception. If there are no adjacent channels, the bandwidth can be increased and thus harmonic distortion reduced.

Automatic interference suppression
A further measure for improving reception is automatic interference suppression, which suppresses interference signals generated by interference sources from both the present vehicle and other vehicles. To this end, the demodulated signal, which contains interference pulses, as well as the useful signal, is blanked for that moment of interference and the gap created is filled.

References
[1] Radio Regulations. International Telecommunication Union (ITU).

Traffic telematics

Transmission paths

Traffic telematics refers to systems that transmit traffic-related information to and from vehicles and, generally, which analyze this information automatically. Both unidirectional broadcasting (radio) and bidirectional mobile communications connections are available as transmission paths. Analog and digital radio provides only the path into the vehicle, with the information received identical for all receivers. Mobile communications connections on the other hand enable messages to be specifically received by way of individual requests and thus information to be exchanged between a vehicle and appropriate service providers in both directions.

The transmission paths also differ in the bandwidth limiting the information volume that is available and the transmission costs. The available bandwidth in mobile communications has increased by several orders of magnitude with each generation in recent years through the developments of GPRS (General Packet Radio Service), UMTS (Universal Mobile Telecommunication System), and finally LTE (Long Term Evolution). While data transmission by radio does not involve additional costs for receiving sound radio, costs are generally incurred for the information volumes on the mobile communications channel in proportion to the data volume.

In the future communication with “roadside units” in accordance with the standard WLAN 802.11p (Wireless Local Area Network) will serve as a further transmission path providing a high bandwidth with no additional transmission costs.

Standardization

The standardization of message content is an important precondition to enable the evaluation of information from various sources by different on-board terminals.

RDS/TMC standard
The widespread RDS/TMC standard (Radio Data System/Traffic Message Channel) for FM radio comprises content regarding the type of traffic disruptions (e.g. traffic jam, full road block), the causes (e.g. accident, icy roads), the expected duration, as well as identification of the road sections concerned. A numeric coding for traffic junctions, expressway sections and geographic regions already exists in many countries. However, it limits the messages for the most part to the major roads (expressways and highways).

TPEG standard
It is now also possible, via the TPEG standard (Transport Protocol Experts Group), to transmit traffic forecasts and recommended alternative routes. A limitation by numeric, predefined junctions and road sections within the road network, as used in TMC, is no longer necessary when the AGORA-C standard is used. This standard allows messages for any road to be coded without transmitters and receivers having to use identical-version reference tables. However, because of the bandwidth required, transmission is only possible via digital channels.

Information recording

The benefit of traffic telematics is dependent on the quality of the messages and how up-to-date they are. There are different data sources for recording information about traffic flow. "Historical" data are additionally used so as also to offer – independently of current messages – the best possible route planning.

Local information
The first data sources for automatically determining the traffic situation were induction loops integrated in the road surfaces which can record the average speed and number of vehicles at individual road points and transmit the data to an evaluating control center. This form of recording was supplemented by the introduction of sensors on expressway bridges which also collect this information.

Area-wide information
In order to receive area-wide information instead of local measurements, newer approaches resort to floating-car data or floating-phone data. The basic idea is to infer traffic situations from the movement data of many cars or mobile phones.

Floating-car principle
In the case of the floating-car principle, the in-car navigation device transmits the car's position and speed by mobile communications to a control center, which then calculates the current traffic situation by statistically analyzing these data.

Floating-phone principle
In the case of the floating-phone principle, the movement patterns of mobile phones are analyzed. This principle exploits the fact that every mobile phone is constantly transmitting information on the current reception situation to its base station. Characteristic patterns can be derived from the changed messages which are accompanied by the mobile communication device's change of location. These patterns are then used to infer the position of the device. Statistical processes are used first to evaluate whether the devices in question are moving devices in vehicles and then to infer the traffic situation from them.

Benefits of the floating processes
The two floating processes are used not only to detect the traffic disruption itself, but also to determine and forward reference values for the delay caused by the disruption.

The movement profiles of the vehicles can also be collected, assigned a time, and compressed. This practice produces so-called hydrographs which reproduce for individual road sections the average driving speeds depending on the day of the week and the time of the day. In this way, recurring traffic disruptions such as slow-moving traffic and regular rush-hour traffic jams are depicted. The vehicle-navigation system can use these hydrographs as "historical" data as a function of the current time to specify a route and journey time adapted to the expected situation.

Dynamic route guidance
Through the standardized coding of traffic messages, it is possible also to display these messages in the vehicle in comprehensible text form for the driver. Furthermore, the computers in route-guidance systems can determined on the basis of the standardized coding whether a better alternative route exists in the event of disruptions. The relevant messages are filtered out from those available using the vehicle location and, if necessary, its movement along a route. If the associated amount of time lost is communicated as well as the actual disruption, this time lost can also be included in the route computation. In the event of alternative route guidance, the driver is instructed that the route has been recalculated on account of traffic announcements received. Further route recommendations will follow in accordance with the new route (see Vehicle navigation).

Furthermore, subscription traffic services not only transmit the more serious traffic jams and disruptions, but also indicate for all higher road categories the

existence of restricted driving speeds identifiable from the floating data. Thus, for example, the traffic situation can be made transparent by a color-coded, high-detail display of the traffic flows in the vehicle navigation's map display even when route guidance is inactive.

Vehicle-vehicle communication

Vehicle-vehicle communication is currently being tested as the logical next step in traffic telematics. The ultimate aim is to further increase road safety and to reduce economic losses due to traffic disruptions. A system of direct information transfer between vehicles can be used to send warnings about, for example, broken-down vehicles, approaching emergency vehicles, the location of the tail end of a jam, right through to relevant individual maneuvers such as hard braking. This gives the driver in the receiving vehicle time to react in an suitable way to the present situation.

The standard WLAN 802.11p (Wireless Local Area Network) serves as the basis for direct vehicle-vehicle communication. The basic principle behind the warning functions is the timely and correct communication of precise location information between the transmitting vehicle and a receiving vehicle. This involves both the continuous transmission of movement messages and the additional transmission of an extra warning message when disruptions and events are detected. The process of regularly repeating or even passing on received messages to other vehicles ensures that all the affected vehicles can be reached.

To prepare for market launch, engineers are currently testing the functionality and efficacy of the systems under everyday conditions in a variety of field tests conducted in for example Germany and the USA.

Driver-assistance systems

Introduction – driver assistance

Driver-assistance systems in modern vehicles can basically be divided into the categories comfort/convenience systems and safety systems. Comfort/convenience systems relieve the driver in performing monotonous, repetitive driving tasks. Typical examples are automatic turn-signal resetting after turning or Adaptive Cruise Control (ACC). Safety systems, on the other hand, are intended to support the driver in critical driving situations so that an accident is avoided or its consequences are lessened. Typical examples are driving-dynamics control (Electronic Stability Control, ESC) and airbags. As the specified system categories increasingly interact in the face of the increasing level of system networking in vehicles and driver support by comfort/convenience systems is intended to provide increased safety (avoidance of critical situations already in advance), there is increasingly a flowing transition between comfort/convenience and safety systems. For further details, refer to the descriptions in the section "Comfort/convenience and safety systems".

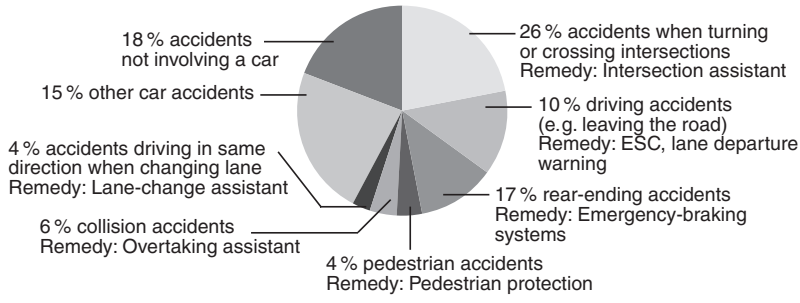
Critical driving situations

Driver-assistance systems aim to make the vehicle capable of perceiving its surroundings, interpret them, identify critical situations, and assist the driver in performing driving maneuvers. The aim is to identify critical situations early, i.e., already before they occur, and anticipate and defuse them, or at best to prevent accidents during critical situations or at worst to minimize as much as possible the consequences of an accident.

In critical driving situations, just fractions of a second are often decisive in determining whether an accident occurs or not. According to the study in [1], approximately 60 % of rear-end collisions and approximately 30 % of frontal impacts could have been avoided if the driver had reacted only half a second earlier. Every second accident at an intersection could be prevented by faster and correct driver reaction.

At the end of the 1980s, when the possibility of highly efficient and partially automated road traffic was presented as part of the EU "Prometheus" project, the electronic components for this task were not in existence. The highly sensitive sensors and powerful microcomputers now available have brought such vehicles driven by highly automated means a step nearer to realization.

Figure 1: Operating areas of passenger-car safety functions in accidents involving personal injury by accident type
(Analysis of GIDAS accident database [2]).



The first driver-assistance systems which facilitate semi-automated vehicle guidance in specific traffic situations have been on the market since 2013 (e.g., traffic-jam assistant). Here the environment sensors scan the vehicle's surroundings and detect relevant objects. The driver-assistance systems follow a cascade of intervention and generate warnings or directly perform the maneuvers required (automatic steering or braking interventions). All this is done at those crucial fractions of a second faster, since a computer-controlled system can essentially react faster than a person.

Accident situation and actions

Accident research plays a significant role in the development of driver-assistance functions. Accident research at Bosch provides support among others in the design and development of new vehicle-safety functions while taking into account current traffic-accident occurrences. Assessments are conducted as to the effectiveness of these systems – i.e., of the potential for accident avoidance – and their effects on future accident occurrences.

To assess the effectiveness of passenger-car functions of driver assistance, an analysis of accidents involving personal injury was conducted for Germany on the basis of the accident database GIDAS (German In-Depth Accident Study [2]) (Figure 1). In Germany in 2015 there were 305,659 accidents involving personal injury, 82 % of which involved a passenger car [3].

Around 26 % of accidents involving personal injury in which a passenger car is involved occur during a turning or crossing maneuver. Here different intersection-assistant functions will be able in future to reduce accidents.

Every tenth accident involving personal injury (10 %) can be put down to driver error; in many such cases involving passenger cars, accidents can already be avoided or the consequences of the accidents lessened with driving-dynamics control (ESC) and lane departure warning.

A further 17 % of all accidents involving personal injury and passenger cars arise when one vehicle rams into another vehicle driving in the same direction. Collision-warning systems can counteract these accidents in an initial stage (e.g., Adaptive Cruise Control). In a further stage collision-avoidance systems prevent the accident by actively intervening in the driving dynamics, for example through brake intervention with the emergency-braking system (Automatic Emergency Braking, AEB).

Accidents involving unprotected road users such as pedestrians or cyclists demonstrate a high degree of complexity. The first emergency-braking pedestrian-protection functions for passenger cars are already available on the market. These operate in up to 4 % of all accidents involving pedestrians. Research is currently being conducted into extending these pedestrian-protection systems to further reduce the occurrence or the consequences of such an accident by means of automatic avoidance maneuvers.

Collision accidents which usually occur during overtaking and can be positively influenced by overtaking-assistance systems amount to 6 %. An additional reduction in accidents by 2 % or in total 4 % can be expected when the driver is supported when changing lanes by an assistance system.

In spite of the already high traffic-safety standards, according to the accident research conducted by Bosch up to 40 % of accidents involving personal injury and passenger cars still do not involve the operation of a vehicle-safety system – accident research is helping here to develop further vehicle-safety functions or to advance their market acceptance.

Applications
Driver-assistance systems have a wide range of applications. They can be divided into active systems which intervene in the driving dynamics and into passive system with no intervention in the driving dynamics.

As already mentioned, a further distinction is made between comfort/convenience systems to reduce the burden on the driver (driver support) with the long-term goal of fully automated driving and safety systems with the aim of accident avoidance or mitigation of accident consequences.

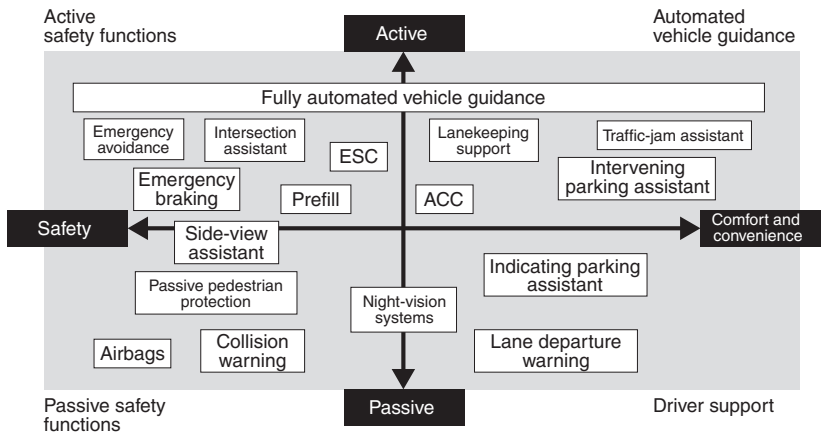
Comfort/convenience and safety systems
Figure 2 shows according to the selected system the range of driver-assistance systems and functions.

Passive safety functions
Passive safety functions (Figure 2, bottom left quadrant) contain the actions for mitigating the consequences of accidents such as airbag deployment and the functions for passive pedestrian protection (specific design of the front of the vehicle to mitigate the consequences of collisions with pedestrians).

Driver support
Systems for reducing the burden on the driver (driver support) without active vehicle intervention (Figure 2, bottom right quadrant) are the preliminary stage to fully automated vehicle guidance. These systems give the driver tips on vehicle guidance. The parking assistant with short-range sensors (ultrasound sensors) aids the driver in looking for a parking space and actually parking, while specific infrared video sensors can be used to good effect to improve driver vision during nighttime driving. In the case of the lane-departure warning system, a video camera uses the road markings to detect the direction of the lane ahead of the vehicle and warns the driver if he/she changes lane without activating a turn signal. The warning can be issued acoustically through the car radio's speakers or mechanically in the form of steering-wheel vibrations.

Automated vehicle guidance
The automated vehicle-guidance systems (Figure 2, top right quadrant) include lane-keeping support, which prevents the vehicle from leaving the lane by means of specific intervention in the lateral guidance. Lane-keeping support is therefore a

Figure 2: Systematic classification of driver-assistance systems (with selected example functions)
Prefill: Brake prefill.



further development of the lane-departure warning system. Adaptive Cruise Control (ACC) is also one of the automated vehicle-guidance systems. A further development of this system relieves the burden on the driver in slow-moving congested traffic – first by braking the vehicle to a complete stop, and then by moving it forward again at low driving speed (ACC Stop & Go). The traffic-jam assistant as a further development also permits lateral guidance at low speeds. Significantly higher requirements are imposed on the functions which are intended to facilitate automated longitudinal and lateral guidance on expressways/highways (highway assistant). The next conceivable development step is full longitudinal and lateral guidance in the urban range, i.e., in unstructured environments, at medium speeds (City ACC). Such a system is the last development stage before highly automated vehicle guidance (see Degree of automation).

Active safety functions

The active safety functions (Figure 2, top left quadrant) concern all the active emergency measures for avoiding or mitigating the damage caused by accidents. In the event of a collision danger that can no longer be averted by the driver, these systems in their maximum configuration enables computer-aided maneuvers to be carried out. Collision prevention or mitigation of the consequences is achieved by automated braking (emergency braking, intersection assistant) or steering (emergency avoidance) here. The intermediate steps consist of a brake prefill when a danger is detected, short hard braking (kinesthetic warning signal to the driver) or a steering pulse. ESC is an example of an active safety system on the vehicle-stabilization level.

Standard architecture for driver-assistance systems

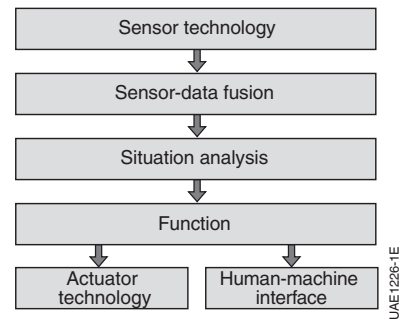
Driver-assistance systems can be subdivided in accordance with a distributed standard architecture (Figure 3) into the submodules sensor technology, sensor-data fusion (with the aim of creating an environment model), situation analysis (with the aim of understanding a traffic situation and its further development), function (with the aim of action planning), actuator technology, and human-machine interaction.

The sensor layer consists of various sensors defined and designed in response to the functional requirements in type and parameterization. In the sensor-data fusion module the environment is modeled in accordance with the function specification (non-moving environment, function-relevant moving objects) on the basis of the sensor measured values.

In the following situation analysis the criticality with regard to function-relevant scenarios is checked on the basis of environment modeling. Thus, the position of the object and the actual position of the vehicle are predicted into the future by way of example for an emergency-braking function and the risk of collision is evaluated.

The criticality measurements derived in this way serve as the criterion in the following function module for a decision on whether to deploy the function. The function module typically uses a finite state machine which as well as thresholds for the criticality measurements also checks whether the function-specific system limits are adhered to (e.g., in the event of the vehicle being subject to high yaw rates an emergency-braking function would not be triggered for the situation where vehicles are driving in the same direction). In the event of a critical situation the function module runs through a cascade of intervention. This ranges, depending on the

Figure 3: Standard architecture for driver-assistance systems



situation and function, from early driver notification and warning (visual, acoustic, haptic) to an automated steering or braking intervention to avoid a collision or mitigate the consequences of an accident. The corresponding actuators and the human-machine interface are now activated in accordance with the function-specific warning cascade.

Typical approaches for the environment-sensor system, sensor-data fusion and function are described in more detail in the following on the basis of the standard architecture introduced.

Environment-sensor system for electronic all-around visibility

Using "electronic all-around visibility", numerous driver-assistance systems are achievable – both for passive and for active intervention purposes. Figure 4 shows the detection ranges of the all-around visibility sensors presented below.

Long range

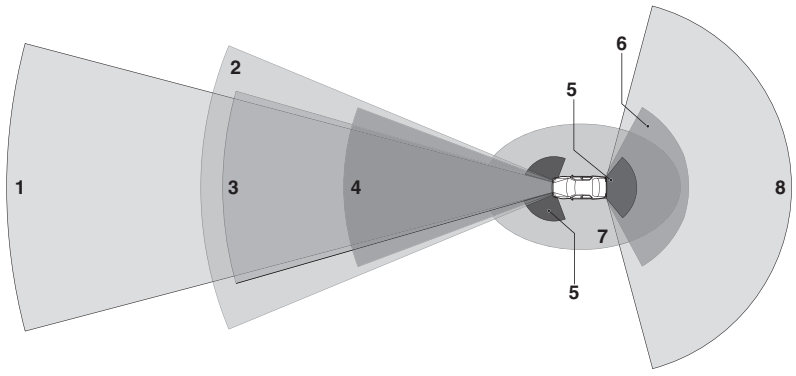
Mainly long-range radar (LRR) sensors are used for long-range applications. Today's sensors have an operating frequency of 76.5 GHz and a range of approximately 250 m. Lidar sensors are also used (e.g., for adaptive cruise control) primarily in Japan. These sensors operating in the close infrared range have ranges of between 150 and 250 m.

Medium range

In the medium range mid-range radar (MRR) sensors with an operating frequency of 24 GHz can meet the exacting demands with regard to range and angle resolution. Since 2005 video sensors have also been used in the medium range. They increasingly play a central role in driver-assistance systems as they enable numerous driver-assistance functions to be implemented due to the large information content of image data (among others thanks to stereo video, optical flow, object classification, see Computer vision) and a very good cost-benefit ratio. For example, infrared cameras in night-vision systems support the driver by detecting situations in the dark.

Figure 4: Detection ranges of sensors for vehicle allround visibility

The detection ranges shown as circle segments indicate the beam angle, however the sensor range decreases as the angles get bigger.
1 Long-range radar up to 250 m, horizontal beam angle 30°,
2 Front-end mid-range radar up to 160 m, horizontal beam angle 45°,
3 Night-vision camera up to 150 m, horizontal beam angle 32°,
4 Video, stereo video up to 80 m, horizontal beam angle 41°,
5 Ultrasonic sensors up to 5 m, horizontal beam angle in each case 120°,
6 Rearview camera up to 15 m, horizontal beam angle 130°,
7 Close-range cameras up to 10 m, beam angle in each case 130°,
8 Rear-end mid-range radar up to 100 m, horizontal beam angle 150°.



Close range
Ultrasonic sensors with higher scanning ranges than previously can form a "virtual safety shield" around the vehicle within certain limits (range smaller than 20 m). This shield can be used to implement a number of functions. The signals from objects inside this safety shield serves as the database of various safety and comfort/convenience systems. Even the driver's "blind spot" can be monitored by these sensors.

Since 2008 multi-beam lidar sensors have been used to monitor the range up to 8 m ahead of the front of the vehicle for automatic emergency-braking functions and have been rapidly taking hold in the volume segment.

Ultra-close range
The parking assistant monitors the ultra-close range with the aid of ultrasound technology. Today's sensors have a range of up to 5 m. These sensors are therefore suitable for a variety of parking functions. Rear-end video sensors (in their simplest form) can assist the ultrasound-based ParkPilot system during parking maneuvers.

Sensor-data fusion
In the event of long-range functional requirements a number of sensors (based on different sensor principles) can be used with an overlapping field of vision in order among other things to increase the field of vision and improve the accuracy and reliability of object detection. This procedure is known as sensor-data fusion (see also Sensor-data fusion). Thus, the camera mounted at the front end can, depending on its functionality, also be used to extend the measured values of the long-range radar. This enables not only the distance to an object to be measured, but also the object itself to be classified. The object class enables the system to draw conclusions about the position and nature of future object movement and thereby helps to increase the reliability of the function. Connecting the video system to the long-range radar creates synergy effects. This significantly increases the beam angle of the ACC system or lane-keeping support and makes object detection even faster and more reliable.

Future driver-assistance systems will incorporate a large number of environment sensors. Because the sensor data typically arrive at different frequencies and asynchronously and are subject to different time delays, a common time allocation must first be established in the fusion module. This is achieved typically using ring memory structures and time stamps. The vehicle's surroundings are modeled on the basis of the sensor data. Grid approaches are increasingly being used to model the non-moving (stationary) surroundings [4]. Moving (dynamic) objects are typically modeled as objects with various function-relevant attributes (such as position, expansion, speed) and time-stabilized by means of an algorithm for state estimation (with Kalman filter [5]). Hybrid approaches which combine object models and stationary grids are still in the research stage [6].

Test and hedging
The primary motivation behind active-intervention systems is increasing driving safety, as is the aim for example with an emergency-braking system (AEB) for automatic collision avoidance.

An important aspect of such systems is the risk of incorrect automatic interventions, for example an unauthorized emergency brake deployment. In the worst case this could result in accidents and consequently the increase in system safety would be undermined. It must therefore be possible through suitable measures to rule out such unauthorized interventions with absolute reliability.

In order to determine the achievable benefit and rule out potential risks, comprehensive tests are therefore part and parcel of the development of new driver-assistance systems. How such tests are conducted and the complexity of such tests currently differ depending on whether the purpose is to test the aimed-for effectiveness in critical driving situations or the hedging against unauthorized interventions.

Because critical driving situations, e.g., the rear-end collisions addressed by AEB, fortunately only occur very rarely in ordinary road traffic, the effectiveness is instead analyzed without endangering

persons on enclosed test sites and with dummy vehicles.

In contrast to these benefit cases, whose parameters can generally be well described from an accident analysis, the hedging against unauthorized interventions must essentially be conducted for all relevant ambient conditions and driving situations. For this reason, the vehicle manufacturers and system suppliers conduct very extensive vehicle tests for the purpose of releasing the system [7].

The main reason for the extensive endurance tests lies in the complexity of the environment sensors used, the wide variety of traffic situations, and the high requirements imposed on the safety of autonomous driving-dynamics interventions. This is because, even in a driving situation that is non-critical for the human driver, "phantom" objects may be detected (e.g., a cloud of exhaust gas incorrectly interpreted by a camera as an obstacle). Because unauthorized braking or steering interventions involve a high risk potential, it is essential to ensure that even very rare cases of such incorrect deployments are discovered before the release of the system. This is the reason for the extensive driving tests.

References

[1] K. Enke: Possibilities for Improving Safety within the Driver Vehicle Environment Loop; 7th International Technical Conference on Experimental Safety Vehicle, Paris (1979).

[2] GIDAS (German In-Depth Accident Study) 2014; http://www.gidas.org.

[3] DESTATIS, Fachserie 8, Reihe 7; Verkehr 2015. Statistisches Bundesamt Wiesbaden, 2016.

[4] C. Coue et al.: Bayesian Occupancy Filtering for Multitarget Tracking: an Automotive Application. The International Journal of Robotics Research, Vol. 25 No. 1, pp. 19–30. HAL archives-ouvertes, 2006.

[5] Y. Bar-Shalom et al.: Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software. Wiley & Sons, New York 2001.

[6] J. Effertz: Autonome Fahrzeugführung in urbaner Umgebung durch Kombination objekt- und kartenbasierter Umfeldmodelle. Institut für Regelungstechnik, TU Braunschweig, Dissertation, 2009.

[7] A. Weitzel et al.: Absicherungsstrategien für Fahrerassistenzsysteme mit Umfeldwahrnehmung. Berichte der Bundesanstalt für Straßenwesen, Heft F98. Wirtschaftsverlag NW, 2014.

Ultrasonic sensor system

Ultrasonic sensor

Application

Today's parking-aid and maneuvering systems (see Parking-aid systems) use ultrasound-technology short-range sensors with a detection range of up to 5.5 m. They are integrated in the bumpers of motor vehicles and serve to calculate distances to obstacles and to monitor the vehicle's surroundings when parking and maneuvering (Figure 1). Sound or even visual alerts are issued to the driver as the vehicle approaches an obstacle.

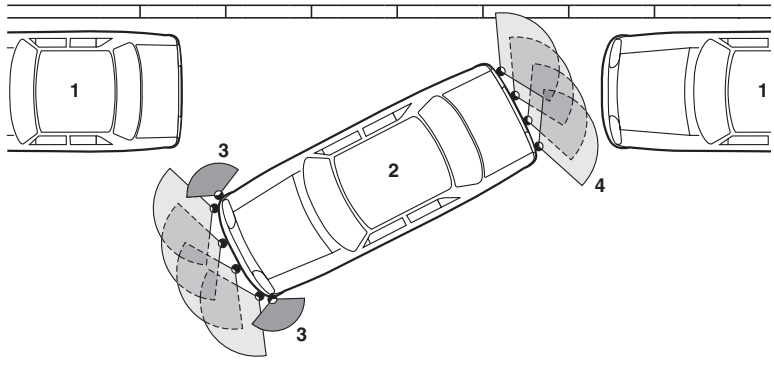
The sixth-generation ultrasonic sensor from Bosch is characterized by digital signal processing, which prevents the systems of different vehicles from influencing each other, and by blindness detection (in the event of icing or fouling) and improved short-range detection. These sensors facilitate comfortable parking in very small parking spaces and maneuvering in tight situations. They also permit the use of a parking-aid assistant, which either issues instructions to the driver for optimum parking or steers the vehicle into the parking space while the driver only needs to take care of longitudinal guidance of the vehicle (see Parking-aid assistant).

Sixth-generation ultrasonic sensors also support emergency-braking functions at low speed by detecting the presence of very near objects and reacting quickly to suddenly appearing obstacles (e.g., pedestrians). In the higher speed range the ultrasonic sensor can also be used to monitor the blind spot (see Blind-spot detection).

Ultrasonic-sensor design

Figure 2 shows by way of example a fourth-generation ultrasonic sensor. It is comprised of a plastic case with integrated plug-in connection, an ultrasonic transducer (aluminum pot with diaphragm onto the inside of which has been glued a piezoceramic element), and a printed-circuit board with transmit and evaluation electronics. They are electrically connected to the ECU (e.g., of a parking aid) by three wires, two of which supply the power. The third, bidirectional wire is responsible for activating the transmit func-

Figure 1: Ultrasonic sensors on a vehicle for use in parking systems
1 Parked vehicles,
2 Parking vehicle,
3 Detection range of the ultrasonic sensors in the front area,
4 Detection range of the ultrasonic sensors in the rear area.



tion and returning the evaluated received signal to the ECU [1].

Ultrasonic-sensor operating principle

The ultrasonic sensor receives a digital transmit pulse from the ECU the parking aid for example. The transmit electronics then induce the aluminum diaphragm to oscillate with square-wave pulses of approximately 300 µs duration at resonant frequency (around 48 kHz) with the result that ultrasonic pulses are emitted. No reception is possible during the relaxation time of approximately 900 µs. The diaphragm, which has meanwhile returned to rest, is made to vibrate again by the sound reflected back from an obstacle. These renewed oscillations are outputted by the piezoceramic element as analog electrical signals and amplified and converted into a digital signal by the sensor electronics (evaluation electronics).

Ultrasonic sensors usually have for the described application a selective emission characteristic (Figure 3) with a wide horizontal sensing range and a narrow vertical sensing range (see Detection characteristics).

Distance measurement with ultrasound

The ultrasonic sensors emit ultrasonic pulses and receive the sound reflected by an obstacle (Figure 4). The piezoceramic element generates an analog electrical signal, which is evaluated in the sensor.

Echo-sounding method

Following the echo-sounding method (Figure 5), the ultrasonic sensors detect the time taken between the ultrasonic

Figure 3: Emission diagram of an ultrasonic sensor

1 Horizontal,
2 Vertical.

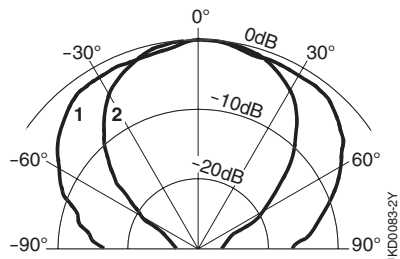


Figure 2: Cutaway view of an ultrasonic sensor

1 Piezoceramic element,
2 Decoupling ring,
3 Plastic case with plug connector,
4 Printed-circuit board with transmit and evaluation electronics,
5 Transformer,
6 Bond wire,
7 Aluminum diaphragm.

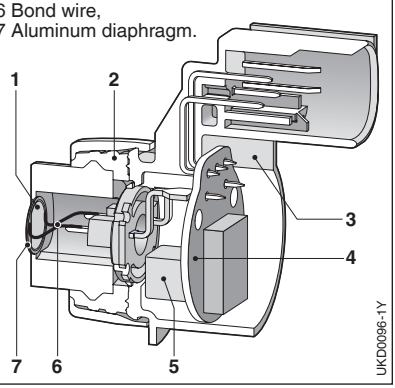
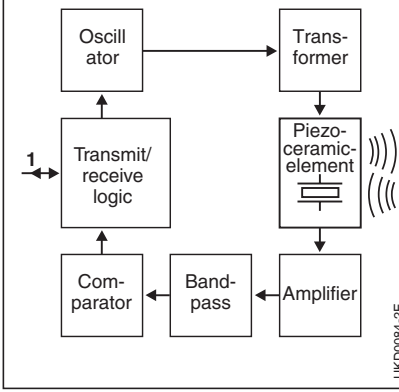


Figure 4: Block diagram of ultrasonic sensor

1 Bidirectional wire to the ECU.



pulses being emitted and the echo pulses being reflected back from obstacles. The distance l between the transceiver probe and the nearest obstacle is derived from the propagation time t_e of the first echo pulses and the velocity of sound c according to the following equation:

$l = 0.5\ t_e\ c$,
where
t_e propagation time of ultrasonic signal,
c velocity of sound in air ($c \approx 340$ m/s).

Detection characteristics
To enable the system to cover as large a spatial area as possible, the detection characteristics must meet special requirements. In the horizontal direction a wide detection angle is desirable to be able to detect as many objects as possible. In the vertical direction, on the other hand, a compromise is required. In order to avoid interference from ground reflections, the detection angle must not be too wide. However, existing obstacles must be reliably detected. Figure 3 shows the emission characteristic of a sensor in the horizontal and vertical planes. A nearly seamless scanning range is produced by the overlapping scanning ranges of mul-

Figure 5: Principle of distance measurement with ultrasound (echo-sounding method)
a) Setup: ultrasonic pulse is output by the emitter, reflected by the obstacle and picked up by the receiver.
b) Signal characteristic.

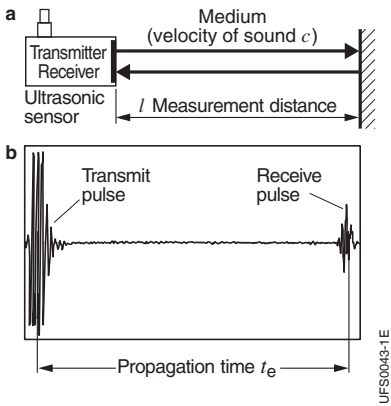


Figure 6: Scanning range of an arrangement with four ultrasonic sensors

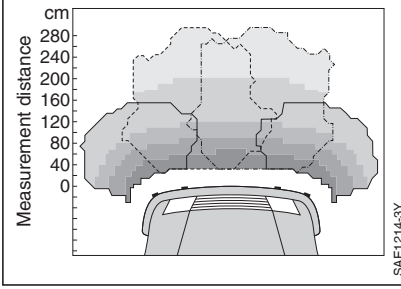
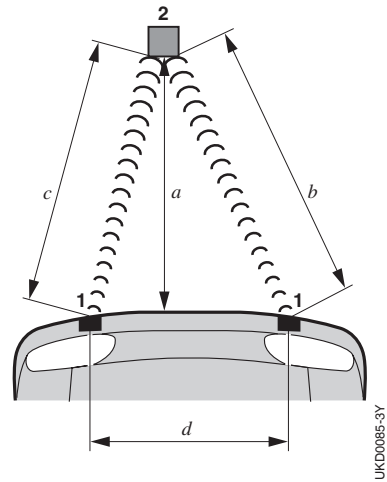


Figure 7: Distance calculation with ultrasound for a simple obstacle
1 Ultrasonic sensors,
2 Obstacle (e.g., post).
a Bumper/obstacle distance,
b Sensor 1/obstacle distance,
c Sensor 2/obstacle distance,
d Sensor 1/sensor 2 distance.



tiple sensors. Figure 6 shows an example of the scanning range of a 4-channel system.

Newer-generation sensors also allow for a variation of the vertical detection angle within certain limits, thus permitting optimized adaptation to the geometry of the bumper and vehicle without resulting false warnings in the event of bumps.

Distance calculation

The geometrical distance a between an obstacle and the front of the vehicle is determined using the triangulation method from the results measured (distances b and c) from two ultrasonic sensors, mounted at a distance d from one another (Figure 7). The distance a is calculated from the following equation:

$$a = \sqrt{c^2 - \frac{(d^2 + c^2 - b^2)^2}{4d^2}} \quad .$$

References

[1] H. Winner; S. Hakuli; F. Lotz; C. Singer (Editors): Handbuch Fahrerassistenzsysteme. 3rd Ed., Verlag Springer Vieweg, 2015.

Radar-sensor system

Application

Radar technology has since 1999 been used with increasing regularity in motor vehicles and supports drivers in their driving tasks. The radar sensor detects objects in front of the driver's vehicle (e.g., other vehicles, pedestrians, roadside structures) and measures their distances, relative speeds and transverse offset to the driver's vehicle. On the road to autonomous driving the radar sensor thus makes a significant contribution to sensing and interpreting the surroundings and the driving situation.

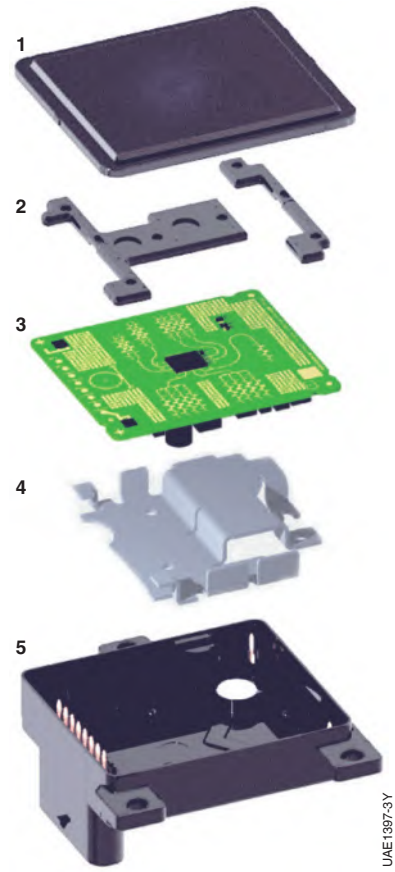
In the early days the radar sensor was used primarily for adaptive cruise control (ACC). This system uses the measurements, adapts the vehicle's driving speed to the traffic in front, and thereby relieves the driver in their driving tasks (see Adaptive cruise control).

The radar sensor's strengths lie primarily in its fast, precise, and weather-independent measurement of distance and relative speed, as a result of which it is particularly suitable for active and passive safety functions. These are for example the Predictive Emergency Braking Systems (PEBS, or Automated Emergency Breaking, AEB), pre-crash detection, lane-change assistant, and blind-spot monitoring. The radar sensor is, aside from the video-sensor system, one of the cornerstones of vehicle-environment sensing on the road to autonomous driving.

Figure 1 shows a radar sensor as a schematic exploded view.

Figure 1: Design of a radar sensor (exploded view)

1 Random (radar-transparent),
2 Absorber (radar-non-transparent, damping),
3 Printed-circuit board with antenna and MMICs (Millimeter-Wave Integrated Circuits, remaining electronics on the underside),
4 Heating-conducting plate,
5 Housing with pressure-equalization element and pins for electrical connection from p.c.b. to vehicle connector.



Radar principle

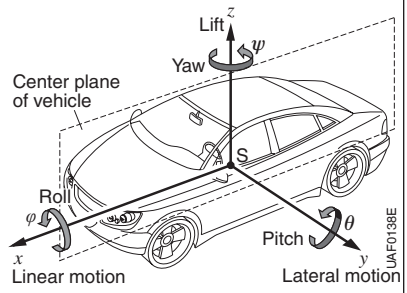
Sensing of objects

The electromagnetic waves emitted by the radar device are rebounded against metal surfaces or other less well reflecting materials and are picked up again by the radar's receiving section. The radial distance from the radar sensor to the objects in the sensing range can be measured from the propagation times of these waves. The Doppler shift of the reflected electromagnetic waves enables the radial relative speed of all the detected objects to be measured directly. The spatial position is determined with the radial distance and one or more angle estimations based on the evaluation of three or more sending and receiving antennas.

Typical objects in the vehicle surroundings generate for the most part a number of reflexes, which depending on the configuration of the radar sensor are also measured separately. The upshot of this is that in the real environment for each measurement cycle several hundred reflexes are detected and forwarded for interpretation to the higher system layers.

Figure 2: System of coordinates

ψ Yaw angle,
φ Roll angle,
θ Pitch angle,
S Center of gravity.



Measured variables in the system of vehicle coordinates

All measured variables are converted to the system of perpendicular and cartesian vehicle coordinates. This is defined according to ISO 8855 [1] as follows: The coordinate datum is usually at the vehicle's center of gravity; the x-axis runs horizontally along the vehicle longitudinal axis and points forwards; the y-axis is vertical to this and points horizontally to the left; the z-axis points upwards (Figure 2).

Technology

Development of radar sensors

Aside from the antenna, the high-frequency electronic components constitute the core of radar sensors. Up to around 2010 the dominant components were GaAs semiconductors (gallium arsenide) for use in high-frequency technology beyond 10 GHz. In the Bosch LRR2 (produced between 2004 and 2013) discrete Gunn diodes made of GaAs were used to generate the 77 GHz carrier frequency. The greatest innovation in the Bosch LRR3 (produced from 2009 onwards) was the switch from discrete high-frequency components to highly integrated MMICs (Millimeter-Wave Integrated Circuits), which have integrated frequency generation, frequency regulation, and also demodulation completely in an SiGe MMIC (silicon-germanium MMIC). This has both reduced the costs dramatically and increased the performance of the radar sensor. The SiGe technology generates on a classic silicon wafer bipolar mixed structures with germanium components which permit very high cut-off frequencies and thus make the silicon-based technology ready for use at 77 GHz. The combination of SiGe technology with classic CMOS technology (BiCMOS technology) on a wafer also permits further high integration for example for signal processing in the same IC.

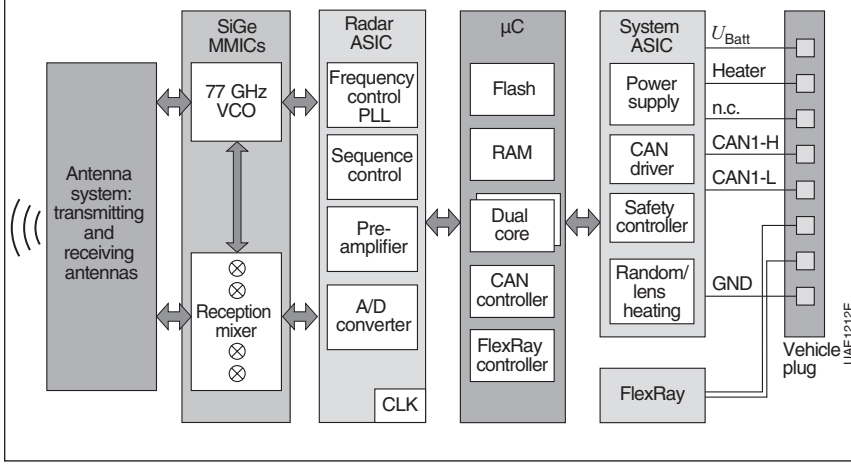
Future development
The increasing reduction of the structure sizes for the pure CMOS processes, which is driven in particular by smartphones, servers, and desktop computers, in the meantime enables the frequencies around 77 GHz to be generated with sufficient efficiency without using SiGe structures. Added to this is the ever more pronounced penetration of the automobile market by radar sensors and the accompanying massive increase in numbers. This now permits the use of the high one-time costs of manufacturing 28 nm, 22 nm or even 16 nm structures for the radar-sensor semiconductors in the automotive sector. The trend towards high integration is thus continued and the future will in part belong to the "single-chip radar" – MMIC together with signal processing and large computer components in an RFCMOS IC (radio-frequency CMOS).

Measurement methods
The received signals are compared with the transmitted signal. Due to the propagation time of the radar waves the received signals are a superposition of the reflected transmitted signals per object and noise created in the reception path. The reflected waves per object are delayed on account of the propagation time and shifted in frequency via the Doppler effect. The transmitted waves are modulated to enable a received signal to be uniquely assigned to a transmitted signal. Whereas previously pulse modulation was used with pulses of a magnitude of 0.5 to 20 ns (corresponding to a distance resolution of 0.15 to 6 m), the predominant method since 2010 has been frequency modulation (also called FMCW, frequency modulated continuous wave), which during transmission alters the instantaneous frequency of the waves over time and requires less hardware outlay.

A further improvement in the detection of objects is made possible by so-called fast FMCW modulation, also known as chirp sequence.

FMCW modulation
Figure 3 shows the block diagram of an FMCW radar. A 77 GHz VCO (voltage-controlled oscillator) supplies the transmitting antenna. The receiving antenna transmits the signal reflected by the object to the reception mixers, which mix the received signal with the current transmitted signal of the VCO and thus transform to lower frequencies in the range from virtually 0 Hz up to several MHz. The high-

Figure 3: Block diagram of a 4-channel FMCW radar
SiGe MMIC: mm-wave-IC, integrated circuit (silicon-germanium technology).



est occurring frequency depends on the maximum propagation time of the signals and the gradient of the frequency ramp. The signals are amplified, digitized, and to determine the frequencies subjected to a high-speed Fourier analysis (see Fourier transform).

Frequency generation
The operating principle of frequency generation is explained in the following. The frequency of the 77 GHz VCO is compared continually by PLL closed-loop control (phase-locked loop) with a stable quartz-based reference oscillator and regulated to a prespecified setpoint value. The PLL is altered during a measurement in such a way that for the transmit frequency f_S a linearly rising frequency ramp is created over time – followed by a linearly falling frequency ramp (Figure 4). The mean transmit frequency is f_0.

Measurement without relative speed
The signal reflected and received from a vehicle driving in front f_E is delayed according to the propagation time, i.e., it has in the rising ramp a lower frequency and in the falling ramp a frequency f_E higher by the same amount Δf_{FMCW}. The amount of the frequency difference Δf_{FMCW} is a direct measure of the distance d, depending on the gradient s of the ramp and the speed of light c:

$$\Delta f_{FMCW} = |f_S - f_E| = \frac{2s}{c} \cdot d \ .$$

Measurement with relative speed
If there is additionally a relative speed Δv to the vehicle traveling in front, the receive frequency f_E is, on account of the Doppler effect, increased (when approaching) or decreased (when the distance increases) in both the rising and falling ramps by a specific, proportional amount Δf_D:

$$\Delta f_D = \frac{2 f_0}{c} \cdot \Delta v \quad \text{(approximation for } v << c\text{)}.$$

In other words, there are two different difference frequencies Δf_1 and Δf_2. For the rising ramp:

$$|\Delta f_1| = |f_S - f_E| = \Delta f_{FMCW} - \Delta f_D$$
$$= \frac{2}{c} \cdot (s\,d - f_0 \Delta v).$$

For the falling ramp:

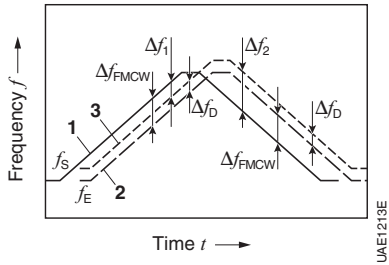
$$|\Delta f_2| = |f_S - f_E| = \Delta f_{FMCW} + \Delta f_D$$
$$= \frac{2}{c} \cdot (s\,d + f_0 \Delta v).$$

Their addition produces the distance d, and their subtraction the relative speed Δv of the objects:

$$d = \frac{c}{4s} \cdot (\Delta f_2 + \Delta f_1),$$

$$\Delta v = \frac{c}{4 f_0} \cdot (\Delta f_2 - \Delta f_1).$$

Figure 4: Distance and speed measurement with a linear FMCW radar
1 Transmit frequency f_S,
2 Receive frequency f_E without relative speed,
3 Receive frequency f_E with relative speed.
f_{FMCW} Frequency difference between transmitted and received radar signals,
f_1 Frequency difference on the rising ramp with relative speed,
f_2 Frequency difference on the falling ramp with relative speed,
f_D Change of receive frequency with relative speed.



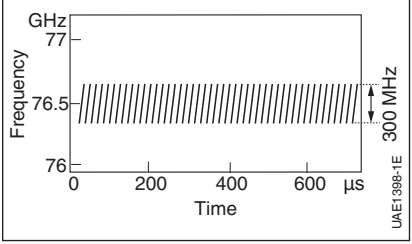
Fast FMCW modulation (chirp sequence)

In the case of fast FMCW modulation many similar FMCW ramps are sent in sequence in a measurement, hence the term chirp sequence. An individual ramp is much shorter with a duration in the region of 10 to 50 µs. This results in a much higher (in terms of amount) ramp gradient such that the amount of the frequency difference Δf_{FMCW} is now much greater than the shift Δf_D due to the Doppler effect; the observed difference frequency Δf is therefore dominated by the distance. For a target distance of 100 m and a gradient of 300 MHz in 10 µs the distance-related frequency shift is 20 MHz. Facing this is a Doppler shift of 41 kHz for a fast target with a relative speed of 80 m/s (= 288 km/h) and a mean transmit frequency of 76.5 GHz.

The fast Fourier transform of the received signals of a ramp produces a discrete frequency spectrum whose frequency cells can be assigned to a distance. These are therefore called distance or range cells.

The Doppler effect is then evaluated by evaluating each distance cell across all ramps, again with fast Fourier transforms. This corresponds to a sampling of the Doppler effect, the individual ramps defined the sampling times and influence the determination of the Doppler frequency. Accordingly the acquired signals are evaluated with a two-dimensional Fourier transform, with the time signal as the matrix of the dimension (number of samples per ramp) × (number of ramps). The two frequencies are accordingly proportional to distance (ignoring the Doppler effect) and relative speed of the object. Figure 5 shows a sequence of FMCW ramps, Figure 6 a two-dimensional spectrum with an object.

Figure 5: Sequence of FMCW ramps



Advantages and disadvantages

The major advantage of this method is the higher number of simultaneously detectable objects and the simpler allocation of distance and relative speed via the two frequencies.

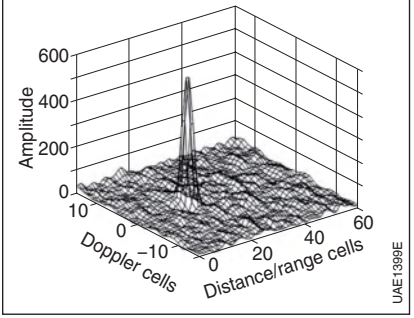
The major disadvantage is the substantial increase in outlay. For the above example due to the Nyquist condition an A/D converter sampling rate of at least 40 MHz is required, 51.2 MHz produces for a ramp duration of 10 µs exactly 512 sampling values. To achieve suitable Doppler acquisition and a higher signal-to-noise ratio, it is necessary to evaluate for example 128 ramps – compared with two ramps a data volume that is higher by a factor of 64. For this reason, hardware units are usually used with calculate the Fourier transforms very quickly and with little power loss.

Antenna system and determining the angle

The antenna system (see Figure 1) has not only to transmit and receive the high-frequency signals, but also to determine the spatial position of the objects. This is required to assign the objects to lanes and to determine object characteristics.

Figure 6: Two-dimensional spectrum with an object

Standardized values at the axes.



Determining the relative position of the objects

Radar systems determine the relative position by estimating the angle under which the receiving antenna locates an object. Two, or better still more than two receiving antennas are required for this purpose. This can be done by directing a single radar beam (scanning) and also by several parallel, overlapping radar beams.

The gains of the complex amplitudes (phases and amounts) which are measured for an object in neighboring beams allow a conclusion to be drawn about the relative reception angle to the radar-sensor axis. In practice four radar beams are frequently used, enabling an angle accuracy of up to 0.1° and a separability of up to 3° to be achieved.

Patch-array-antenna system

There are many ways of technically realizing the antenna system. The most common variant is the patch-array-antenna system with antenna elements which are realized in microstrip technology.

Design

The patch-array-antenna system, in contrast to the lens-antenna system, as a rule has a bistatic design. The transmitting antenna in this system consists of a large number of single patches arranged in rows and columns which are wired in such a way that superposition produces a concentrated radar beam along similar lines to concentration by a lens.

The receiving antenna typically consists of a number of single patches arranged in columns which are electrically separated from each other and arranged next to each other. The offset of the individual reception columns generates a phase shift between the received signals of a detected object which is used to determine the angle. A reception column can in an individual case also consist of a number of wired columns.

Determining the angle

Angle evaluation utilizes the dependence of the antenna characteristic in amount and phase on the direction to the object of the outgoing or incident waves. To evaluate the outgoing waves a number of transmitting antennas are required which exhibit a directionality compared to its characteristic, for example via different directions of maximum radiation (this results in different amounts, often used for monopulse methods) or via different positions (this results in different path lengths to the object and thus in phase differences, often used in “phased array”). A number of receiving antennas can be used and evaluated in the same way.

If both a number of transmitting antennas and a number of receiving antennas are used for angle evaluation, the term used, based on communication technology, is MIMO systems (multiple input multiple output). This enables the direction to the object to be determined even more accurately.

Variants of radar sensors

In the past a distinction was made between short-range radars (SRR, usually in the 24 GHz frequency band) and long-range radars (LRR, in the 76 GHz frequency band). In the meantime the applications are merging and sensors are increasingly covering both ranges such that mid-range radars (MRR, in the 76 to 81 GHz band) will dominate the market, supplemented by a long-range radar. Short-range radars have a typical range of 20 to 50 m and a beam angle of up to 160°. Long-range radars have ranges of approximately 300 m and beam angles of up to 100°; mid-range radars have ranges and beam angles in between.

Vehicles in the future will often have five or more radar sensors all round the outer skin, placing a virtually continuous envelope around the vehicle and providing the opportunity for further assistance functions to make driving safer and more comfortable.

Homologation

Radar sensors are in accordance with the requirements devices which transmit and receive radio waves. Their operation is therefore governed by country-specific regulations worldwide. So-called homologation ensures that for each country in which radar sensors are to be operated the legal requirements are satisfied and the sensor receives type approval.

Radar sensors are usually approved by the manufacturer, e.g., Robert Bosch GmbH. The vehicle manufacturers then refer in the Owner's Manual or on specially set-up websites to type approval by the manufacturer.

Homologation covers both the technical aspects (such as frequency range, transmitted power, directional characteristics) and the labeling of radar sensors. Using the EU as an example, a radar sensor must among other things carry a specifically sized CE mark in a visible location on the sensor. Homologation must both be completed before the radar sensor is put on the market for the first time and, depending on the country, also be renewed on a regular basis.

References

[1] ISO 8855: Road vehicles – Vehicle dynamics and road-holding ability – Vocabulary.

Lidar-sensor system

Functions and applications

Previous applications
The use of lidar sensors (lidar, Light Detection And Ranging) for applications in motor vehicles began in Japan and in the USA around the year 2000, initially for adaptive cruise control (ACC). Since 2008 a low-priced multibeam lidar has tapped into the mass market when used for a rudimentary, automatic emergency-braking function. Since 2018 a multilevel laser scanner for an expressway/freeway traffic-jam assistant and an improved emergency-braking function have been in series production (for the first time including pedestrian protection as per the NCAP standard; NCAP, New Car Assessment Programme, [1]).

Future applications
Since 2010 lidar sensors with constantly increasing resolution and range have been in development. Above all the development of powerful lasers and low-noise, high-resolution image receptors have led to the situation today where comparatively cheap solutions are available. By 2025 – driven by functional safety – a dramatic increase in the number of lidar sensors for highly automated driving functions is expected (SAE levels 4 and 5, see Future of automated driving). Derived from this is the need for a redundant sensor design with wherever possible different physical measurement principles. In other words, lidar sensors will in future closely accompany video and radar sensors. Examples of the first driving functions in development with lidar are the urban automated taxi and the first expressway/freeway pilots in the truck and passenger-car sectors.

Operating principle

Distance measurement
Lidar sensors operate with laser beams in the infrared range (IR). A light beam emitted by a transmitting module is reflected by an object and detected by a receiving module. The time of flight between transmission and reception produces the distance between the object and the lidar sensor (see Measurement principles).

Intensity measurement
As well as the time of flight the intensity (power) of the reflected light can also be measured. The amount of reflected power essentially depends on the distance to the object and its IR reflectivity (see Time-of-flight principle, Power balance). It is therefore possible using additional model assumptions to extrapolate from intensity to reflectivity, an additional object characteristic.

Image build-up
Distance measurement is used to measure a distance value between the lidar sensor and an object located at the solid angle captured by the light (e.g., distance to the vehicle traveling in front).

To resolve the surroundings spatially as a three-dimensional image, it is necessary to capture many solid angles within a desired field of vision independently of each other. There are two known fundamental principles for this purpose. Firstly, the number of transmitting elements or receiving elements can be increased by further, spatially independent pairs. Within a pair there can be transmitting and receiving elements in any ratios (see Methods for increasing the number of pixels, Single-beam, Multibeam, Flash lidar). Secondly, the light beams can be deflected via a scanner in a horizontal or vertical direction (Figure 1, see also Methods for increasing the number of pixels, Macro-, Micro-scanner).

Depending on the distance and resolution, the captured objects must be detected in great detail in a high-resolution 3D image. In this way, contours of

vehicles, persons, and other objects, including in particular the road surface and infrastructure, can be detected in great detail. At short distances to the sensor even body language and hand gestures of persons can be resolved.

The lidar sensor captures an intensity image as well as the 3D image. Unlike a video camera, however, it contains only light from a narrow IR wavelength range (only "one color"). Depending on the wavelength range, such intensity images can be similar to the capture in a video camera or the human eye, but also differ significantly. Furthermore, retro-reflectors (e.g., rear reflectors) appear unusually intensive due to the active measurement principle. The intensity resolution (image dynamics) is much lower than a video camera.

Environment perception (perception)

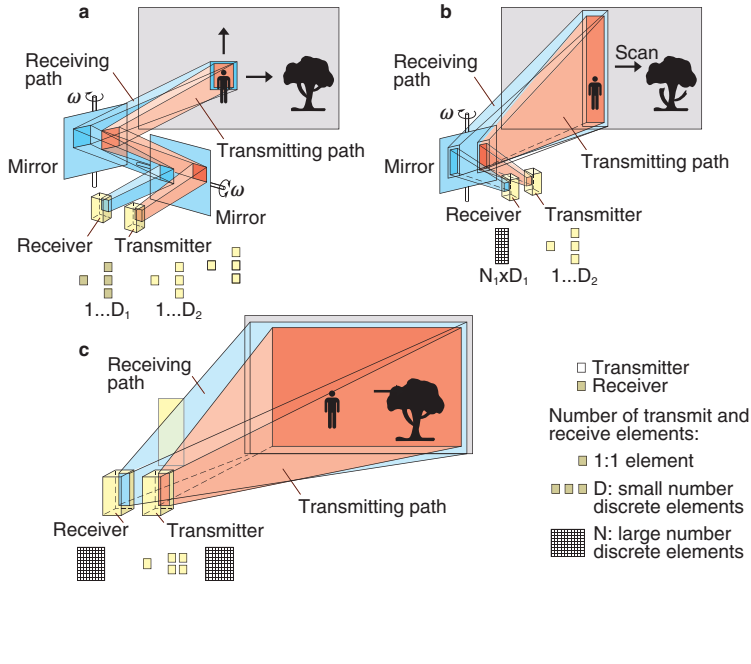
Lidar sensors supply on the detection level a large number of 3D pixels, also called a 3D point cloud, which are for the most part enriched by additional attributes (typically time stamp, pulse intensity, pulse length).

More abstract environment descriptions can be created using software algorithms on the basis of this 3D point cloud. Typical features on the object level for lidar sensors:

– objects (dynamics, 3D contour),
– lanes,
– open spaces,
– roadside structures,
– object classification (type, dynamics, object height, and derived from this ability to drive over small obstacles on the road).

Figure 1: Design variations of lidar sensors

a) Spot scan with 2D deflection,
b) Vertical flash with 1D deflection,
c) Full flash.



The lidar is additionally used to
– estimate proper motion,
– locate the vehicle within the area around the vehicle,
– have the area around the vehicle mapped by map providers.

Additional and monitoring functions
For use in highly automated driving functions sensors must in future adhere not only to rigorous requirements relating to functional safety (functional safety, focus on E/E faults), but also relating to very high availability of the intended function (SOTIF, Safety of the Intended Function).

It is possible to derive from functional safety and SOTIF a series of additional and monitoring functions which the lidar sensor itself can conduct. Required monitoring functions are:
– detection of contamination,
– detection of dazzle/glare,
– estimation of weather conditions,
– estimation of visibility and range,
– estimation of extrinsic calibration (i.e., of the time-varying installation parameters).

Required additional functions are:
– control of a heating system (to de-ice the optical window),
– control of a cleaning system (to clean the optical window),
– control of a cooling system.

Advantages of lidar
The lidar sensor has significant advantages over the video camera:
– high distance accuracy even over long distances (200 to 300 m),
– full range of functions at night or in unlit parts of the surroundings,
– capture of all object geometries (mono camera: only trained object classes; stereo camera: only sufficiently large object surfaces with texture),
– capture and classification of small objects, e.g., shed load and its height and resulting ability to drive over it.

It has significant advantages over radar:
– high angle accuracy over all distances (radar: angle uncertainty on stretched objects; objects with incorrect angle, caused by interference),
– detection of objects with virtually all geometries, sizes, and surface properties (radar: reflectivity sometimes too low depending on view angle, size, material),
– classification of objects based on contour capture (radar: object contour cannot be seamlessly captured),
– detection of lanes (radar: no detection of road surfaces and lanes).

Disadvantages of lidar
Lidar sensor has some disadvantages compared to the video camera:
– lower image resolution (number of distinguishable object classes is lower),
– lower intensity resolution (detection of classification based on texture, e.g., characteristic object classes, texts, only possible with video camera),
– intensity contains only a small proportion of the color spectrum (identical IR gray-scale value for non-identical color).

It has significant disadvantages compared to radar:
– lower robustness against adverse weather conditions, particularly fog, sometimes spray (but not mist/haze, moderate rain, snowfall),
– direct glare from the sun (essentially only at the solid angle of the sun),
– less accurate speed measurement (this is done only by way of one measurement, but is derived in terms of time from the position change).

Properties and use

Ranges
Comparable to other environment sensors in the motor vehicle, lidar sensors can also be classified according to their typical detection range. The following details of ranges always relate to weakly reflecting objects (worst case). The maximum range can typically be taken as double the value.

Near-field lidar
Near-field lidar has a range of almost 0 m to around 30 m and ideally a field of view approaching a hemisphere (180° horizontally and 180° vertically). The angle resolution is typically 0.5 to 2°. If a number of such lidar sensors are distributed around the vehicle, uninterrupted monitoring of the near field is possible. This facilitates for example highly automated parking and safe lane changing on expressways/freeways. However, due to the availability of other low-priced alternative sensors (video, radar, ultrasonic), there are as yet no known equivalent lidar developments.

Mid-range lidar
Mid-range lidar has a range of around 0.5 m to 80 m and, depending on the application, a field of view of around 50 to 150° horizontally and 10 to 20° vertically. The angle resolution is typically 0.15 to 0.25° horizontally, 0.15 to 0.85° vertically. A sensor mounted on the front of the vehicle supports semi-automated ranging systems (expressway/freeway assistant, traffic-jam assistant) and emergency-braking functions, based on radar and camera. Highly automated driving functions can only be realized with them in a very limited speed range and in very simple driving scenarios. For semi-automated driving functions mid-range lidar has been in series production (at Denso) for over 20 years in small and fluctuating quantities. Since a further manufacturer (Valeo) arrived on the market in 2017, numerous manufacturers have been announcing their arrival on the market with larger quantities for the period between 2021 and 2023.

Long-range lidar
Long-range lidar has a range of around 1 to 150 m and a field of view comparable to mid-range lidar. The angle resolution is horizontally and vertically around 0.08 to 0.15°. With three to six sensors distributed around the vehicle highly automated driving functions in the urban environment are possible. A front sensor (if necessary extended to include a rear sensor) facilitates highly automated driving functions on expressways/freeways. Here lidar is on an equal footing with radar and camera. A number of development projects since 2010 having been focusing on this lidar class, which is among the fundamental components of the highly automated driving functions.

Ultra-long-range lidar
Ultra-long-range lidar has a range of around 200 to 300 m and a field of view of 50 to 90°. The angle of resolution is horizontally and vertically below 0.05°. A sensor mounted at the front permits high driving speeds well above 100 km/h and full monitoring of large intersection areas. Lidar bears the main responsibility for these tasks compared to radar and camera.

Even if the development activities, apart from one exception, are in their infancy, the first sensors will be ready to go into production by 2023.

Properties
Distance accuracy
Some properties are shared by all lidar classes to a large extent. Thus distance accuracy registers for the most part between 5 and 15 cm. The frame rate registers for the most part between 10 and 25 Hz.

Image resolution
The image resolution is very different, ranging between some 1,000 and several 100,000 pixels. Here each pixel contains as a 3D pixel at least one piece of distance information and one piece of direction information, frequently one piece of intensity information (comparable to the infrared image of an infrared camera), less frequently additional classification information.

Disadvantageous properties
The most important disadvantageous properties of lidar sensors are today their size (depending on the principle and range up to two liters volume) and their still high cost.

Figure 2: Distance measurement via time of flight
t_1 Transmission time of lidar signal,
t_2 Reception time of reflected signal,
r Distance between lidar sensor and object.

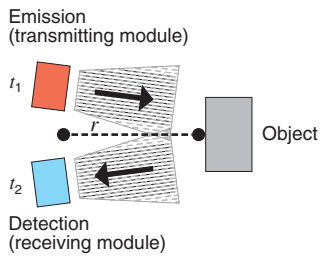
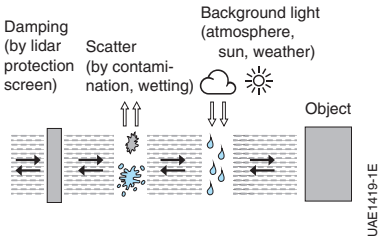


Figure 3: Light losses due to scatter and damping (absorption)



Measurement principles

Time-of-flight principle
Distance measurement via time of flight
The fundamental principle of every lidar sensor is the time-of-flight principle (TOF). A modulated light beam is emitted at a time t_1 by the transmitting module and scattered on an object (Figure 2). Some of the scattered light falls at the later time t_2 back into the receiving module. The time of flight ($t_F = t_2 - t_1$) is proportional to double the distance r between the transmit-receive unit and the object. The distance is calculated via the speed of light c_0 in air at

$$r = 1/2\ c_0\, t_F .$$

Power balance
The lidar equation descriptions the relationship between the light emitted by the lidar and the received utilizable light. The receiving power $P(2r)$ is dependent on six factors:
– the transmitting power P_0,
– the reflectivity R_O of the object,
– the optical losses T_S in the transmitting path,
– the optical losses T_E in the receiving path,
– the receiving aperture D, and
– the distance r.

The following applies:

$$P(2r) = P_0 R_O T_S T_E D^2/(2r)^2$$

If the object surface (horizontal or vertical) is smaller than the diameter of the transmitted beam, the power drops even more sharply with the distance. Thus:

$$P(2r) \sim 1/r^4.$$

Significant losses occur due to scatter and absorption through optical elements, contamination, and the atmosphere (Figure 3). Fog, mist/haze, and spray can result in a substantial scatter of the light. The atmospheric losses are taken into consideration in the optical losses in the transmitting and receiving paths.

Dynamic range
The dynamic range of the receiving power $P(2r)$ is calculated at constant transmitting power P_0 from the reflectivity of the objects and the distance range to be measured at around 100 to 160 dB.

To adapt to the large dynamic range required, some receiving circuits operate primarily in saturation, with the result that only edge detection is possible. In other words, the very high gain of the analog receiving power leads to an almost digital switching performance. Other systems perform a dynamic adjustment of the receiving-power gain, e.g., increasingly with rising time of flight. An adjustment of the transmitting power can also be made.

Signal-to-noise ratio
To safely identify the useful signal, its power must be clearly above all the noise outputs in the system; the signal-to-noise ratio (SNR) must be large enough. Aside from all the electronic noise sources the Sun and artificial light sources constitute a significant noise source.

A widely used method to increase the signal-to-noise ratio is statistical improvement by accumulation of several individual measurements. This accumulation can take place both directly in the detector (PMD, Photonic Mixing Device; SPAD, Single Photon Avalanche Diode) and also in the digital domain.

Wavelength and eye safety
Few wavelength ranges are well suited for lidar sensors. These are derived from the availability of suitable transmitting and receiving components, from eye safety (in accordance with DIN EN 80625-1, [2]), invisibility to humans, and damping by the atmosphere. Important factors for eye safety are wavelength, beam shape, and modulation.

Essentially all popular lidar sensors operated in the infrared range (IR). The demand for a light source that is invisible to humans leads to a wavelength above 850 to 880 nm.

A first atmospheric window then extends up to around 930 nm. 905 nm represents a good compromise between low-priced lasers and silicon detectors whose quantum efficiency decreases rapidly above 850 nm. The limits of the maximum permissible radiation to comply with eye safety are at 905 nm still really low because infrared light close to visible light still strikes the retina virtually undamped. Nevertheless, this is the most widespread wavelength today.

A second atmospheric window for which lasers are available is at 1,300 nm, a third at 1,550 nm. For both wavelengths it is necessary to use other detector substrates (InGaAs, InP) whose maturity, costs, and noise performance are comparatively unfavorable. The limits of eye safety are however much higher because infrared light is to a large extent absorbed by the aqueous humor. The advantage becomes a disadvantage in the detection of wet objects or some objects with a high water content. In all, system performance at these wavelengths has not been examined sufficiently and there is no lidar of this type in series production.

Modulation methods
The signal-to-noise ratio is very much dependent on the type of modulation. Because noise increases with the measurement period, pulse modulation for lidar sensors particularly with medium and high range is the type of modulation favored in practice.

In all direct and indirect pulse time-of-flight methods the amplitude of the light is modulated. Similarly to radar, FMCW modulation methods in which the wavelength is modulated will in future be used for lidar (see FMCW modulation).

Direct pulse time-of-flight method
With the direct pulse time-of-flight method (Figure 4) a comparatively strong, short pulse of light is emitted (typically up to 100 W pulse height, 2 to 15 ns pulse duration, pulse repetition rate up to 200 kHz). The pulse of light reflected by the object (minimum power a few 10 nW) is identified by means of a pulse-detection method and its chronological position and frequently also the signal shape are determined.

Background light is only received in a small time window. The distance is calculated from the chronological position of the received pulse relative to the emitted pulse via the speed of light.

One particularity of the direct pulse time-of-flight method is that even a number of received pulses can be determined in succession. Objects can also then still be detected by means of this so-called multi-target capability (per pixel) if previously a less relevant detection was already determined (e.g., due to noise, spray, fog, reflection, glass).

The reflectivity can be inferred from the signal shape. But a classification can also be made as to for example whether the reflection comes from a hard target (object) or a soft target (atmospheric interference, e.g., fog). Sometimes the signal shape is also used to improve the distance estimation (e.g., walk-error compensation; walk error is an apparent time-of-flight error which occurs for different receiving power levels during a threshold-value comparison). Only the direct pulse time-of-flight method permits a signal classification to determine object properties, contamination or atmospheric interference (Figure 5).

Pulse detection

Two methods are customary for pulse detection. Technically more complicated, but more favorable (lower-loss) with regard to the information content is digital sampling of the received signal with an A/D converter (Figure 6a). The sampled signal can be specifically searched after detections, particularly by means of a correlation analysis (matched filter).

Technically simpler, but involving greater information loss is pulse-edge detection (Figure 6b), realized by a threshold-value comparison that is high-resolution in terms of time (TDC, Time-to-Digital Converter).

The lidar block diagram differs for both pulse-detection methods in one component only (Figure 7a). All the components operate in the high-frequency range (around 1 GHz).

Figure 5: Signal classification with the direct pulse time-of-flight method

1 Object size > beam diameter,
2 Object size > beam diameter,
3 Damping and scatter,
4 Fog and smoke.
r Distance between lidar sensor and object,
α Absorption coefficient of air.

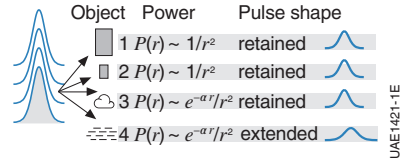
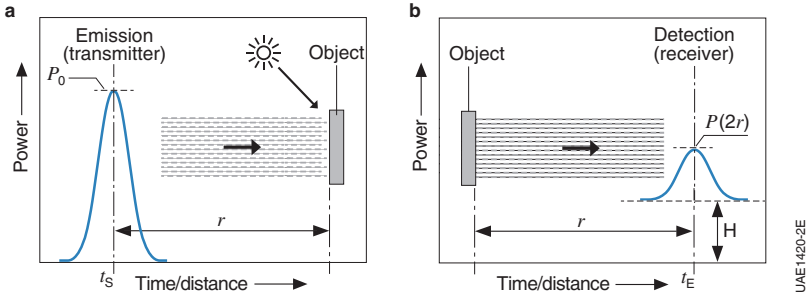


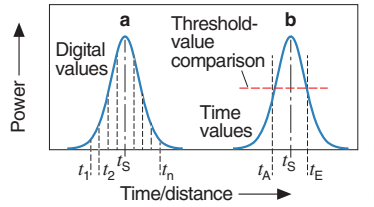
Figure 4: Direct pulse time-of-flight method

a) Transmission of the lidar signal at time t_1,
b) Reception of the reflected signal at time t_2.
r Distance between lidar sensor and object,
P_0 Transmitting power,
$P(2r)$ Receiving power,
t_S Transmission time of lidar signal,
t_E Reception time of reflected signal,
H Background light.



Indirect pulse time-of-flight method
In the case of the indirect pulse time-of-flight method (Figure 7b) a long, wherever possible square-wave pulse with high amplitude is emitted (a few 100 ns pulse duration, 100 W to 100 kW pulse height). The receive pulse is integrated in the receiver in two exposure sections. The contain the phase relation and thus the time of flight between the transmitted and received pulses. This produces the time of flight.

Figure 6: Pulse detection by means of analog-to-digital conversion
a) Analog-to-digital converter,
b) Time-to-digital converter (pulse-edge detection).
$t_1 \ldots t_n$ Sampling values,
t_A Start time, threshold-value crossing,
t_E End time, threshold-value crossing,
t_S Calculated mid-point (center).



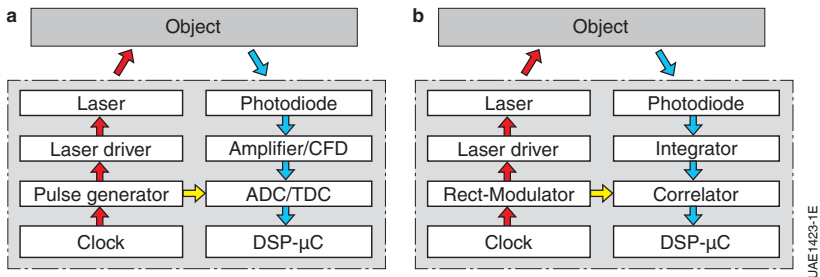
The first exposure section corresponds to the transmission duration so that as the distance to the object increases an ever-decreasing proportion of the reflected light still shines into the gate window and consequently can be integrated (Figure 8, S1). This component falls linearly with the distance of the object. If the object distance is above the distance of the corresponding transmission duration (upper distance limit), no more light can be integrated. This first partial measurement is still ambiguous with regard to object reflectivity and object distance.

The reflectivity of the object can be determined with the second exposure section. This component rises linearly with the distance of the object (Figure 8, S2). Its time duration corresponds to at least twice the transmission duration so that the entire reflected pulse is always recorded.

Both partial measurements in total contain the information about the reflectivity of the object and also an offset by the background light. The background light is therefore determined in a third partial measurement without pulse lighting and subtracted from the integrated value.

The requirements imposed on the dynamic range of the receiver are particularly high for this technology because in addition to the utilizable light the background light is also integrated over a rela-

Figure 7: Lidar block diagram
a) Block diagram for direct pulse time-of-flight method with analog-to-digital converter or time-to-digital converter,
b) Block diagram for indirect pulse time-of-flight method.
ADC Analog-to-digital converter,
TDC Time-to-digital converter,
DSP Digital signal processor,
CFD Computational Fluid Dynamics,
Rect Square-wave function.



tively long period of time whose intensity is much lower. The signal-to-noise ratio is lowered by the additional background light. In all, no systems with very high ranges can be realized today with this method.

Realization variants

There are three realization variants. In the first variant two quickly switchable CMOS imagers (see Imager) are used to carry out one of the two partial measurements each. The received light is distributed via a beam splitter to the two imagers. The distance images are calculated in a downstream stage on a processor. Because the imager is based to a large extent on standardized camera imagers, very high resolutions up to several megapixels can be achieved. The drawback is that incorrect measurements occur due to the time interval between the two partial measurements if the object under consideration performs a rapid movement (movement artefacts).

The second variant implements the principle of beam splitting and distribution to two imagers by means of polarization. A Pockels cell is used instead of two rapidly switchable imagers to rapidly switch the light. The Pockels cell in the process in a few 10 to 100 ns turns the polarization direction of the received light through 90°. The two imagers are each provided with a polarization filter, both filters being turned through 90° to each other. By rotating the polarization direction, one imager receives with the rotation an increasing proportion of the light output from near objects, while the other imager receives that from far objects. The object distance can be calculated from the ratio of the two light outputs.

In the third variant the measurement principle is implemented directly in a special CMOS imager in which each pixel consists of two taps (sensor ranges). One tap takes on one of the two partial measurements. The level of sophistication of these imagers is still comparatively low in terms of pixel size, quantum efficiency, and noise. The more complex circuits and lines on the imager reduce the space factor and consequently give rise to comparatively low resolutions (some 1,000 to 10,000 pixels).

The high pulse power of the light source and the homogeneous light distribution today pose a major technical challenge for all indirect pulse time-of-flight methods and owing to the high costs are currently not competitive.

Figure 8: Indirect pulse time-of-flight method

r Distance between lidar sensor and object,
c Speed of light,
H Background light.
t_1, t_3 Start, end time, transmission pulse,
t_2, t_4 Start, end time, reception pulse,
Δt Transmission-pulse length,
Δt_1, Δt_2 Amount of reception pulse inside/outside the transmission window.

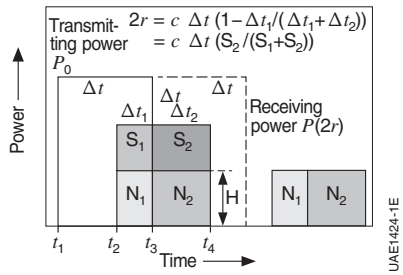
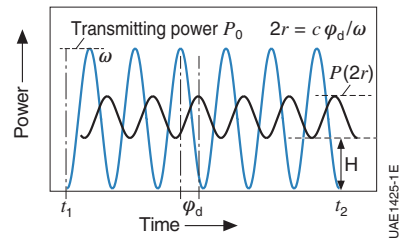


Figure 9: CW phase measurement method

r Distance between lidar sensor and object,
c Speed of light,
φ_d Phase displacement,
ω Modulation frequency of transmitting power,
H Background light.



CW phase measurement method

In the CW phase measurement method (CW, Continuous Wave) the light is continuously emitted and sinusoidally modulated in amplitude. The received signal is by φ_d out of phase (Figure 9). The phase displacement corresponds to the time of flight and thus the distance to the object. The background light is received in a small time window.

The intensity of the emitted light can be taken as sinusoidal. To determine the received signal in full, it is necessary to determine the amplitude, the phase relation, and the offset. Usually this is carried out by integrating the received signals via four phase windows corresponding to the relative phase displacements 0°, 90°, 180°, and 270°.

Central elements of the CW phase measurement method are CW modulation and correlation with the received signal (Figure 10). The received signal is processed at low frequency. Sampling takes place due to the fact that the light converted into charge carriers is integrated according to its relative phase relation in four charge-storage elements. The charge-storage elements are activated and deactivated over a number of measuring cycles. The three unknown variables phase relation (distance), amplitude (intensity), and offset (background light) can then be determined from the charge ratios of the four partial measurements.

Four voltage measurements are conducted using two successive measurements and two taps (per pixel) (Figure 11). Each measurement integrates the charge carriers via a phase of 180°, the two taps being in phase opposition (charge-carrier swing). The two measurements are in quadrature. The phase relation and thus the distance can be determined from the voltage values.

The calculation of the phase relation can already be implemented in the imager. The four partial measurements are either simultaneously realized in four taps of one pixel or in two partial measurements with two taps per pixel. The rapidly switchable taps of each pixel are also known as charge-carrier swing or photonic mixing device (PMD).

To prevent the pixels from being saturated by strong background light, direct components can be removed in part (SBI, Suppression of Background Illumination).

The emitted light output can be limited to 1 to 100 W by the integrating method. The duration of a measurement is some ten milliseconds. The modulation frequency is typically 10 to 20 MHz.

Significant drawbacks are the lack of multi-target capability, the low immunity to interference, and the insufficient dynamics of the imagers. For applications in motor vehicles the development of PMD cameras was driven on by some manufacturers until around 2015 and then to a large extent stopped.

Figure 10: Block diagram of CW phase measurement method

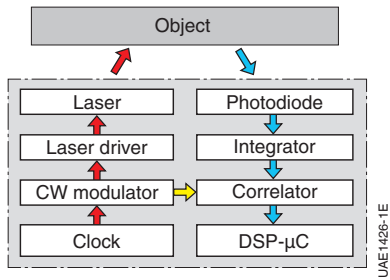
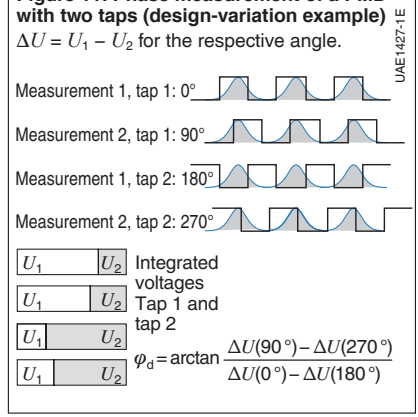


Figure 11: Phase measurement of a PMD with two taps (design-variation example)

$\Delta U = U_1 - U_2$ for the respective angle.



FMCW measurement method
In the FMCW measurement method the wavelength of the emitted light is changed (e.g., ramp). The received light is optically interfered with the transmitted light and consequently the light frequency (≈ 200 to 350 THz) is downmixed into the baseband. Frequency lines in the baseband correspond to the object detections, to which as well as the distance a speed can also be assigned (see FMCW radar).

The usual mechanical approaches can be chosen for beam deflection. Research is also being conducted in electronic beam deflection.

Methods for increasing the number of pixels

Basic principles
If the number of spatially resolved pixels is to be increased, more time-of-flight measurements must be carried out in different directions accordingly. The individual measurements can take place either chronologically in parallel or iteratively, i.e., in succession. Receiving arrays are customary for chronologically parallel measurements. In chronologically successive measurements the beam is deflected in different directions (scanning). A deflector is required for this purpose.

The speed of light is a physical limit on the number of iterative measurements. If frame rate f and maximum range r_{max} are specified, this produces the maximum number of individual measurements

$$N_{max} = c_0 / f\, r_{max}.$$

Example: For 10 Hz image rate and 300 m maximum range 100,000 pixels are obtained. Without further parallelization this amounts to the maximum number of independently measurable pixels.

To realize high-resolution lidar sensors with high range, it is therefore necessary to combine both principles of iterative and parallel measurement.

Solid-state lidar sensors
Single-beam, multibeam, and flash lidar belong to the class of solid-state lidar sensors which operate without mechanically moving components. Figure 12 shows combination options for partitioned transmitting and receiving elements in a solid-state lidar. The number of pixels increases with the number of partitions.

Single-beam lidar (single fixed beam)
In the simplest variant the transmitted beam is focused by a transmitting optic on the desired detection range (i.e., solid angle) and a corresponding receiving optic directs the collected light into an individual receiving element (see Figure 12). Thus no spatially separable distance measurements are possible inside the solid angle. Only one 3D pixel per measuring cycle is generated.

Multibeam lidar (multi fixed beam)
By duplicating components of single-beam lidar it is possible accordingly to measure numerous solid angles separately (see Figure 12). A high degree of duplication has a detrimental effect on costs compared to more highly integrated systems and consequently such systems for the most part only measure three to five solid angles. It is only advisable to use such systems in applications where rough volume elements are to be checked for the fundamental presence of an object. Multibeam lidar is used very successfully for a rudimentary but effective emergency-braking system.

Flash lidar
In this variant of highly integrated, high-resolution lidar the transmitted beam is widened as for single-beam lidar to the entire detection range. The detector consists of a high-resolution imager (FPA, focal plane array, up to 256 x 256 pixels). Each pixel of the imager is able to perform a time-of-flight measurement (Figure 12).

Long ranges in conjunction with large fields of view and resolutions above some 10,000 pixels lead to comparatively high costs and reach the limits of eye safety much earlier. There are also disadvantages with regard to spatial separability.

Macro-scanner with rotor
In this variant a single-beam, multibeam or flash/part-flash lidar is placed on a motor-driven, rotating platform and rotated about the rotor axis (rotor scanner, Figure 13). In the case of flash lidar the transmitted beam is widened to the entire detection range. In contrast, in the case of part-flash lidar, the transmitted beam only covers part of the detection range such that for complete coverage an additional scanner (see Macro-scanner with deflecting mirrors) or an electronic scan is required. The electronic scan is performed through sequential activation of laser lines or laser columns of a laser array (VCSEL, vertical-cavity surface-emitting laser).

The vertical resolution is achieved by a vertical partitioning of transmitting or receiving elements or both. The number of transmitting and receiving elements determines the vertical resolution. The individual measurements can take place chronologically in parallel or iteratively. The horizontal resolution is achieved by rotation. The measurements must accordingly take place iteratively. In total a resolution of several 100,000 3D pixels can be achieved.

Transmitting and receiving modules can be laterally or vertically arranged in a rotor scanner (Figure 13). The number of transmitting and receiving elements determines the vertical resolution.

Advantages and disadvantages
Disadvantages are the higher costs of the mechanical drive, the additional wireless power and data transfer to the rotor, and heat-dissipation challenges. On the other hand, this variant has the advantage of

Figure 12: Combination options for partitioned transmitting and receiving elements in a solid-state lidar

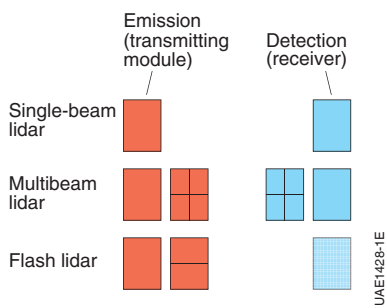
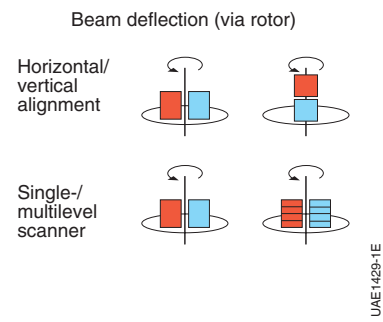


Figure 13: Rotor scanner



a horizontal field of view up to 360° and very homogeneous performance over the entire field of view.

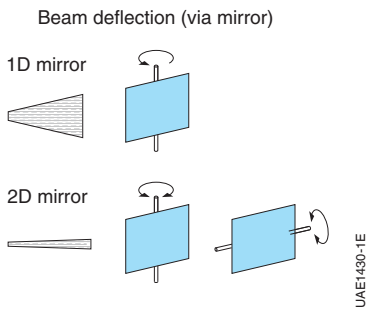
Macro-scanner with deflecting mirrors
In this variant the transmitting and receiving beams of the single-beam, multibeam or part-flash lidar are directed via one or more moving mirrors. In view of the many possibilities two typical system configurations will be described.

In the first configuration the beams of vertically partitioned transmitting and receiving elements are deflected via a rotating mirror that rotates about a vertical axis. The horizontal field of view is limited by the size of the mirror and the vignetting of the housing to 90 to 120°. The resolution is several 100,000 3D pixels.

In the second configuration the beams of individual or few transmitting and receiving elements are directed via two oscillating mirrors with an orthogonal oscillating direction. The beam thus follows the established Lissajous figures. The number of possible pixels is reduced by around 40 % compared to a rotating scanner.

Figure 14 shows two typical configurations of a mirror scanner. A vertically extended beam is directed via a rotating mirror (1D mirror) and only horizontally deflected. In the second configuration a focused beam is directed via two oscillating mirrors (2D mirrors) and deflected in both principal directions.

Figure 14: Configurations of a mirror scanner



Micro-scanner
To miniaturize oscillating mirrors 1D or 2D MEMS micro-mirrors (Micro-Electro-Mechanical Systems) have been analyzed in the last few years. For long ranges these small mirrors (usually under 1 cm diameter) have proven to be an “optical eye of a needle”, which could only be brought under control with the aid of highly voluminous and highly complex optics. In all, micro-mirrors offer no advantages for the currently pursued approaches. But they could play a role again in the context of FMCW lidar.

References
[1] https://www.euroncap.com/de
[2] DIN EN 80625-1: Safety of laser products – Part 1: Equipment classification and requirements.

Video sensors

Application

People use their sense of sight to discern the information pertaining to environment detection. Consequently it is natural to develop driver-assistance systems on the basis of video data. There are now a multitude of assistance functions and the demands on the video cameras also change, depending on the application.

Indicating display systems
The first applications were indicating display systems, such as for example rear-view cameras. In these systems importance is attached to high-contrast images and color reproduction. The images are often reprocessed to obtain the desired impression in the display.

Image-processing systems
As well as indicating display systems there are systems which use computer-based image-data processing (computer vision) to extract relevant objects and determine the driving situation. It is no longer necessary for the images to be perceived to be realistic by people; instead, they must depict relevant objects for the algorithms so that they are easily distinguishable.

Cameras for image recognition can be used for purely informative systems, as with road-sign recognition. But cameras also serve to control actuators, such as
– for automatic headlamp dimming,
– for brake interventions for vehicle-to-vehicle ranging and for accident avoidance,
– for steering interventions, e.g., in parking functions, lane-keeping support or evasive steering support,
– or as sensors for automated or autonomous driving.

The higher the degree of automation, the greater the demands on the performance of the cameras. This is achieved through the interaction of the elements of a camera described in the following. The two main components are the image sensor and the objective (see Figure 1).

Figure 1: Video camera for use in driver-assistance systems
a) Schematic representation of a video camera,
b) Design of a Bosch video camera.
1 Objective, 2 Objective holder (front housing section),
3 Board with image sensor, 4 Rear housing section with connector.

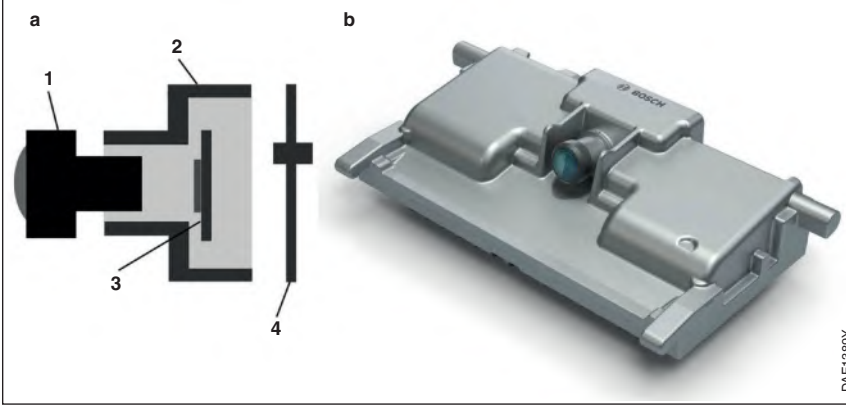


Image sensor

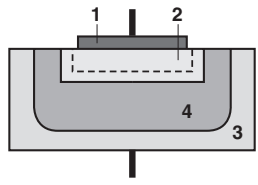
Basic principles of photosensing

Aside from a few special applications, image sensors (imagers) are today manufactured using CMOS technology (complementary metal-oxide semiconductor). This technology enables transistors to be embedded in every pixel for control purposes.

In the image sensor many photodiodes are arranged in a rectangular pixel grid. A layer of p- and a layer of n-conducting semiconductor material meet in a photodiode (Figure 2). An electric field is created at the junction between the two layers, the space-charge region. At the same time this space-charge region demonstrates a certain capacitance which behaves in inverse proportion to its thickness.

Photodiodes typically work by being charged to a certain potential (output voltage) and then being exposed to light. When photons shine into the semiconductor, electron-hole pairs are generated. Here, the quantum efficiency η describes how many electron-hole pairs are generated by one photon. The photoelectric current is proportional over many factors of ten to the incident luminous flux and is linear over a wide dynamic range. The photoelectrically generated charges now discharge the photodiode capacitor. The residual voltage is measured once the photodiode has been exposed to light. The difference between this voltage and the output voltage is a measure of the quantity of incident light.

Figure 2: Photodiode
1 Contact, 2 n silicon substrate,
3 p silicon substrate, 4 Space-charge region.



If so many photons shine in that the space-charge region is completely discharged, further photons no longer lead to a linear reduction of the voltage and the photodiode is saturated. Efforts are usually made to operate the photodiodes in their linear range. This can be achieved under bright lighting by reading out the voltage already after a short exposure time and by coupling the photodiode to an additional capacitor. However, large capacitors provide a lower voltage drop per photoelectron such that dark scenes and the associated low voltage drops are more difficult to measure.

Design

Photodiodes and metal-oxide capacitors can be manufactured on a chip with standard semiconductor processes like CMOS. Large photodiodes have a large surface for collecting photons, meaning that they also exhibit low shot noise in dark scenes. Shot noise arises because only integer electrons can be measured. Stochastically a pair of photons more or less meets the photodiode in any exposure. Depending on the quantum efficiency only a part of the photons is converted into electrons that can be read out. Both processes lead to fluctuations even with homogeneous lighting viewed macroscopically.

The ratio of light-sensitive volume to lateral face is high for large photodiodes such that cross-talk between adjacent photodiodes is easier to handle than for small photodiodes.

Large photodiodes use a lot of silicon surface area, impacting on the sensor costs. The pixels are therefore for the most part designed to be as small as possible to obtain a still acceptable dark sensitivity. The trend towards higher sensor resolutions, i.e., more pixels, provides for ever larger sensors in spite of ever smaller individual pixels. Where in 2010 sensors with a quarter megapixel (VGA resolution) and 5 µm pixels were customary, in 2020 sensors with 2 megapixels of 3 µm each or 8 megapixels of 2.1 µm each are the norm.

Activating the image sensor

Aside from the photodiode, each pixel also includes transistors to control the image sensor. These transistors can for example charge the photodiode to the output voltage are direct the residual voltage to the readout electronics. The setup of a simple pixel with three electrodes is shown in Figure 3. Modern pixel designs often contain further transistors to obtain additional functions or improved behavior.

The control electrodes of the transistors (gate) are connected to each other in lines and therefore can only be switched jointly (Figure 4). The readout lines are connected to each other in columns. If a voltage is applied to the control line to read out a line, all the pixels in this line are simultaneously connected to a readout line.

Each column has its own readout electronics; an entire line can therefore be evaluated in parallel. The voltage is initially analog-amplified. The gain is adjustable for many image sensors and influences the sensor's noise performance.

Evaluating the sensor signal

The pre-amplified signal is then converted in an analog-digital converter (A/D converter) into a number of fixed bit depth. If the analog voltage is above the reference voltage for the highest digital value, a saturation occurs in the digital output. The saturation threshold of the A/D converter is selected in so that the photodiode still operates in its linear range.

A/D converters with high bit depth can show higher dynamics between the brightest and darkest value, but are costlier to produce and are sometimes prone to higher noise. Through the noise contributions of all the analog components only a limited dynamic range can be sensibly used from a single readout.

As at 2020 converters with 10 and 12 bits are widely used in series applications.

Figure 3: Connection of a pixel in the CMOS image sensor with three transistors
PD Photodiode,
M1, M2, M3 Transistors.

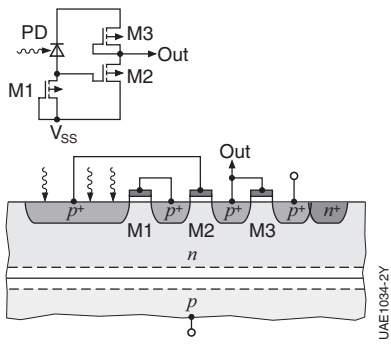
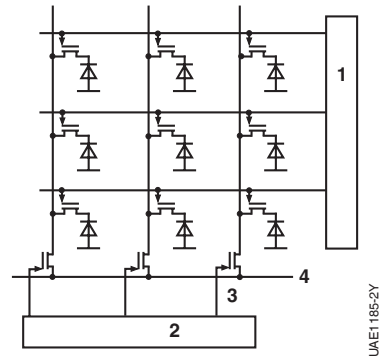


Figure 4: CMOS image sensor
Photodiode array with one photodiode and one activation transistor per pixel.
1 Line addressing,
2 Column addressing,
3 Column selection,
4 Column signal line.



Images with high dynamics
Additional techniques are often used in order also to be able to depict scenes with higher dynamics (high dynamic range, HDR) with low noise and without over- or underexposed areas.

Dual conversion gain
One possibility here is "dual conversion gain". A second capacitor can be connected via a resistor to the photodiode for this purpose. The resistors can be realized for example by means of a half-closed transistor. While the photodiode discharges immediately when exposed, the voltage follows the supplementary capacitor with a delay. At the end of the exposure the photodiode and the supplementary capacitor are read out separately. The small capacitor of the photodiode exhibits already with low luminous fluxes a high voltage drop and is suitable for dark image areas. The large capacitor is saturated much later and can be used for bright image areas. For the most part the digital values of both readouts (capture) are already offset against each other on the image sensor.

Split-pixel technology
If the dynamics are still insufficient, a sensor with split-pixel technology can be used. Here photodiodes with high and low sensitivity are arranged in pairs and again offset against each other. Usually the sensitive pixel is much larger and absorbs through its surface area more photons than its insensitive counterpart.

Multiple exposure with different exposure times
To further increase the dynamics, it is also possible to record in succession and offset exposures of different lengths. In moving scenes, however, artifacts may appear in the form of multiple edges.

Exposure modes
Rolling-shutter mode
Automotive image sensors are for the most part operated in rolling-shutter mode. Here a count variable points at the line currently to be processed. This counter is incremented at a fixed rate, the line rate. Initially, the current line is charged to the output voltage (reset). Then the pixels accumulate the photons and discharge. In a fixed line spacing after the reset line the pixels are connected to the evaluation electronics and converted into digital brightness values. The exposure therefore "rolls" over the sensor.

The advantage of this method is the simple and favorable implementability. The disadvantage is that each line was exposed with different start and end times. Objects may appear distorted as a result.

Global-shutter mode
In global-shutter mode the entire sensor is simultaneously charged to the output voltage. Each photodiode is connected to a light-protected capacitor (shielded cap) belonging to the pixel. At the end of the exposure time all the capacitors are simultaneously disconnected from their photodiodes. The voltages can then be read out in succession.

The advantage of this operating mode is that all the pixels are exposed simultaneously and no motion artifacts appear. It is more complex and therefore more expensive to process shielded caps.

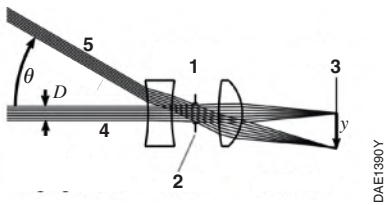
Color filters
To be able to distinguish colors, photodiodes with different color sensitivity must be arranged on the sensor. In practically all applications photodiodes are damped in a regular pattern with different color filters. This pattern is also called a mosaic. The Bayer pattern is the most widely used pattern in photography. It consists of repeating lines of 2×2 pixels. A green filter is applied to two diagonally opposed pixels; the other two pixels are sensitive in the red and the blue (see Figure 9).

Human color perception is likewise based on receptors for red, green and blue. As a result, the camera perceives color spectra similarly to the eye and can reproduce realistic images.

As well as the Bayer pattern there are other color patterns, which can contain for example colorless pixels. Colorless pixels are sensitive to all colors, therefore collect more photons, and have above all in dark scenes advantages in the signal-to-noise ratio. The same applies to magenta, yellow and turquoise (cyan), which transmit in each case two of the three additive primary colors red, green, blue.

If a sensor is offered with a specific number of pixels, this number refers to the sum total of all pixels. There are therefore far fewer pixels available per color channel. The color channels are for the most part interpolated in image preprocessing (see Demosaicing) in order to have available at each location a brightness value for all color channels.

Figure 5: Objective
1 Objective,
2 Screen,
3 Image,
4 Beam bundle along the optical axis,
5 Beam bundle with field angle θ to optical axis.
θ Field angle,
y Image height,
D Diameter of beam bundle.



Objective

The objective reproduces light from an object on the image sensor. In the case of a sharp reproduction a point-shaped light source is focussed on a dot or as small a spot as possible on the sensor. The object-side angle of the incident beam bundle to the optical axis is called the field angle, the distance between the pixel and the optical axis the image height (see Figure 5). A reproduction is characterized by among others the following properties:

Field of view
The field of view is the maximum object-side angle at which an object can still in order to still be reproduced. Objectives are designed for a maximum angle up to which they reproduced sharply. Furthermore, the sharpness and brightness of the reproduction often diminish dramatically and the reproduction can no longer be used. Only areas which are reproduced both within this maximum angle and within the sensor size are reproduced in the camera.

Sensors are rectangular, i.e., the size differs in the horizontal, vertical, and diagonal directions. Connected with this are different fields of view in these directions.

Projection model
The projection model describes how object-side field angles are translated into image heights. Two important projections are the pinhole-camera model and the equidistant projection model.

Pinhole-camera model
The eponymous pinhole camera consists of only an aperture and a film plate (Figure 6a). A pinhole camera translates straight edges into straight lines in the image. Real objectives with lenses are designed in such a way that the carry out the same translation of field of vision into image height as a pinhole camera. The reproduction function is given by

$$y = f \tan \theta.$$

Here y is the image height on the sensor, f the focal length, and θ the field angle. The focal length can be understood in the formula as the enlargement ratio. It should be chosen in the design of the objective in such a way that the desired field of view is reproduced on the sensor. The derivation of the reproduction function, i.e., the change of the image height per angle, is called enlargement. It is often given in the unit pixels per degree. In pinhole cameras the enlargement increases as the field angle increases and diverges at ±90 °; a sensor of infinite size would be necessary. It is therefore clear that real wide-angle objectives must deviate from the pinhole-camera model and follow a different projection model.

Equidistant projection model
Wide-angle and fisheye cameras often use the equidistant projection model (also called the F-theta model). As an example it is possible to image an equidistant camera as a modified pinhole camera with a curved image (Figure 6b). Here the image height is given by

$$y = f\,\theta.$$

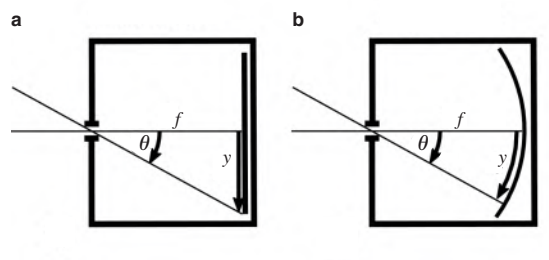
The angle θ must be applied in rad. In this model identical angle changes are translated into identical changes of the image position. The enlargement is constant. Straight lines are generally shown with a barrel distortion.

Distortion
Because many other parameters must also be taken into account in the design of an objective aside from the projection, real objectives rarely follow one of the projection models exactly. The relative deviation of the projection of an objective from a model is called distortion. If no projection model is specified, the pinhole-camera model is assumed for the most part as the reference. Through the clever design of the distortion it is possible to place a high level of distortion in image areas in which a degree of camera performance is required.

Luminous intensity
The luminous intensity of an objective is traditionally described by the f-number familiar from photography. The f-number is determined with a parallel (collimated) beam bundle on the optical axis, which fills the aperture entirely. The f-number is the ratio of focal length to the diameter of the beam bundle (see Figure 5). The smaller the f-number, the greater the luminous intensity of the objective.

Relative illuminance
In most cameras the illuminance decreases with increasing field angle. The ratio of brightness on the sensor at the optical axis to brightness in the field is called relative illuminance. This is caused for example by an angle-dependent increase of reflections or by cutting off part of the beam bundle by means of lens edges (vignetting). The distortion also affects the relative illuminance. A change of enlargement also changes the object-side solid angle from which a pixel collects light.

Figure 6: Projection models
a) Pinhole-camera model,
b) Equidistant projection model.
f Focal length,
y Image height on the sensor,
θ Field angle.



Sharpness

A measure of sharpness widely used in technical optics is the modulation transfer function (MTF). The MTF describes the loss of contrast (Michelson contrast) incurred by a periodic object with a sinusoidal brightness characteristic in the reproduction. A sharp objective can itself reproduce fine structures with high contrast whereas a blurred (unsharp) objective already creates noticeable contrast losses with rough structures. The characteristic period in which the MTF must be measured can vary from task to task: Should for instance only the outer shape of a road sign be recognized or text on the sign?

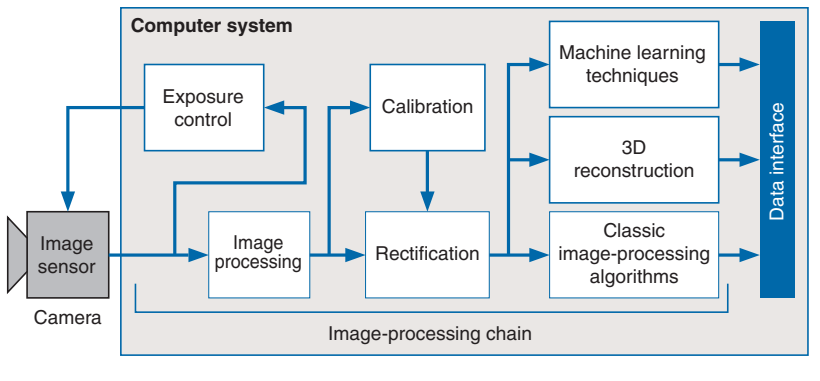
Computer vision

Image-processing chain

The term computer vision (CV) comprises the processing of an image-data sequence by a computer system. The object of image-data processing is to detect and classify objects which are relevant to a specific application. In the field of driver-assistance systems intelligent camera systems are used in which the captured image data must be processed in real time and only the limited computing resources of an electronic system embedded in the vehicle are available.

The block diagram (Figure 7) shows the schematic representation of computer vision for a driver-assistance system. The individual processing steps in their entirely are called the image-processing chain. A digital representation of the environment captured in the range of vision of a camera is created with the aid of the image sensor (imager). This image representation is the database of computer vision and thus forms the input data stream of the image-processing chain. The first processing steps comprise the algorithms of image processing and exposure control fed back to the image sensor. Calibration supplies as a supporting algorithm static and dynamic parameters for rectification. Rectification is followed by the parallel processing of the classic image-processing algorithms, the 3D reconstruction, and the machine learn-

Figure 7: Image-processing chain



ing techniques. To finish, the results of the image-processing chain are transferred via a data interface to the following processing steps on the computer system.

Exposure control

The exposure time of the image sensor is dynamically adapted to the light conditions of the captured scene with the aid of exposure control. In this way, a sufficiently high range of contrast can be guaranteed at any timed in the image data generated by the image sensor. To this end, exposure control uses brightness information from the currently captured image data to optimize the exposure time for the subsequent capture. The image-data stream generated by the image sensor has a frequency of typically 30 Hz or 30 fps (frames per second).

In camera-based driver-assistance systems an HDR image sensor (high dynamic range) is usually used in which differently exposed captures are fused into an overall image with a higher dynamic range. The simplest case here is the combination of an overexposed capture with an underexposed capture. This operation, known as HDR recombination, happens either directly on the image sensor or as part of the subsequent image processing.

Scenes with rapidly changing brightness values such as tunnel entrances and tunnel exits pose a challenge for exposure control.

Image processing

The algorithms of image processing are based on the image data received by the image sensor, which are also called raw data. The typical component parts of image processing, also known as “image signal processing”, are shown in Figure 8

Image processing comprises the following blocks: demosaicing, noise reduction, edge preservation, tone mapping, and color-space conversion. Aside from these blocks there are further processing steps which are typically dependent on the image sensor used, the available computing resources, and the available computing unit.

Demosaicing

An image sensor has a color filter to be able to represent the color information of a captured scene. Each pixel in the image sensor's raw data therefore represents a color value in digital form. An RGGB filter, also known as a Bayer filter, is typically used as a color filter (color filter array). A red, green and blue component is interpolated in the RGB color space for each individual pixel (Figure 9) by demosaicing with the aid of the neighboring pixels. There is generally no reduction of the image resolution by the demosaicing. Numerous publications with algorithms of

Figure 9: RGGB color filter and demosaicing

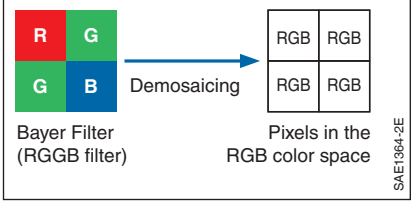
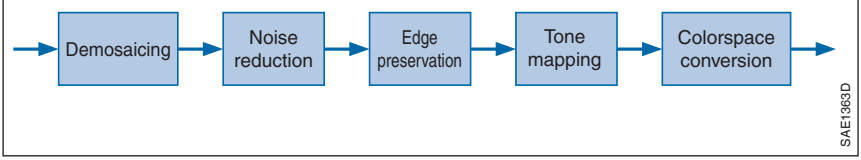


Figure 8: Component parts of image processing (image signal processing)



all manner of complexity are known in the literature to calculate the RGB values [1].

Aside from the RGGB filter, other color filters which typically contain pixels without color information are used for image sensors in camera-based driver-assistance systems. These intensity pixels (clear pixels) are more light-transmitting than color pixels and therefore produce an altogether higher image brightness and as a result a higher-contrast reproduction of dark scenes.

Typical color filters with intensity pixels in the field of driver assistance are for example RCCC (i.e., one pixel with red filter and three intensity pixels) and RCCG (i.e., one pixel with red filter, one pixel with green filter, and two intensity pixels).

Noise reduction

Noise reduction serves to suppress image noise and thus to reduce unwanted signal components, improving the signal-to-noise ratio of the image-data stream. Two-dimensional filter operations are generally used for noise reduction and applied for each pixel. A typical example of a filter for noise reduction is a Gaussian filter [1]. Small image structures like noise for example are suppressed by the use of the smoothing filter. From a signal-theory standpoint the Gaussian filter thus equates to a low-pass filter.

Edge preservation

The image data were smoothed by the filter operations of the previous processing steps. An algorithm for edge preservation is used in order nevertheless to obtain a detailed image with clearly discernible and sharp edge information. There are numerous established techniques for preserving edge information; these techniques are mostly based on filter operations which ideally sharpen only the edges, but influence or smooth the remaining image data as little as possible. A typical example of an edge-preserving smooth filter is a bilateral filter [1].

Tone mapping

The adaptation of contracts within the image data is referred to as tone mapping. This operation can for example increase the range of contrast in dark image areas to facilitate the recognition of existing details there. There are all manner of established global and local algorithms for tone mapping [1]. In most techniques the color value of a pixel is replaced with the aid of a reproduction rule by a different color value. For global techniques a location-independent reproduction rule is used for the full image. In the often more complex local techniques the reproduction rule is dependent on the position of a pixel.

Color-space conversion

In the case of color-space conversion the image information in the RGB color space is carried over for each pixel to a different color space because other color spaces are better suited for some computer-vision algorithms. For example, the YUV color space facilitates through the separation of brightness information (luminance Y) and color information (chrominance U and V) a simplified processing of algorithms which only process luminance values.

Typically, algorithmic techniques for 3D reconstruction only work on the brightness information of an image. Machine learning techniques are typically in a position to work on the basis of information from both the RGB color space and the YUV color space. For color-space conversion there are very many different and in part also internationally standardized conversion rules [1].

Calibration
The calibration of a camera system is required to describe the relationship between a point in the real world and its projection on the focal plane.

Static calibration
The theoretical basis for static calibration is an idealized camera model which describes the mathematical relationship between a point in the real world and its projection on the focal plane. The simplest form of a camera model is the pinhole-camera model. In this model the captured scene is projected through an idealized pinhole without perspective distortion onto the focal plane (Figure 10). The idealized needle hole is also called the pin diaphragm. The pinhole-camera model is typically used for a front-camera system with an aperture angle of less than 120°. The equidistant projection model (fisheye-camera model) is generally used for a rearview camera with a larger aperture angle of more than 180°.

To describe the relationship between real world coordinates and image coordinates, the pinhole-camera model has parameters which are determined during calibration with the aid of pairs of points and are available for coordinate transformation. The connection between coordinates in the world coordinate system and image coordinate system is established with the aid of the camera coordinate system (Figure 11).

Intrinsic parameters
The relationship between camera coordinate system and image coordinate system is described by the intrinsic parameters of the camera model. These parameters include the focal length and the position of the optical center of the camera. On the simplified assumption that the camera only has one single thin lens, the optical center describes the point behind the focal plane in which all the beams of a parallel incident beam bundle cross. The intrinsic parameters are thus camera-specific and describe the internal geometry

Figure 10: Pinhole-camera model

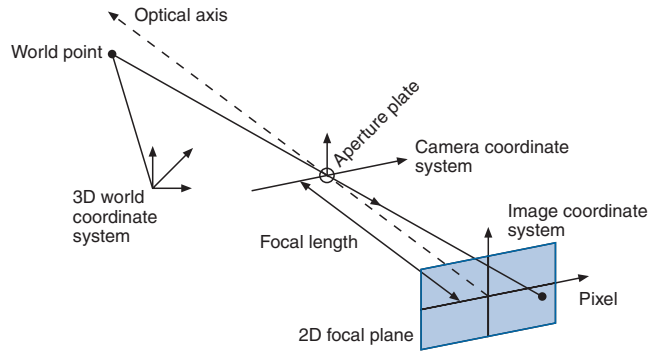
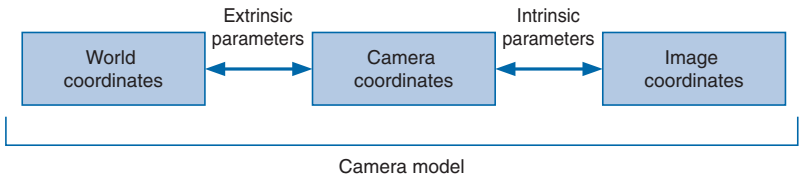


Figure 11: Relationship between real world coordinates and image coordinates



of a camera. The intrinsic parameters are typically determined after the production of a camera and stored in a non-volatile memory inside the camera.

Extrinsic parameters
The relationship between world coordinate system and camera coordinate system is described by the extrinsic parameters of the camera model. A translation and rotation of world coordinates using the extrinsic parameters results in the corresponding camera coordinates. Consequently, the position and orientation of the camera can be described in three-dimensional space via the extrinsic parameters. The extrinsic parameters are determined after the camera is installed in the vehicle because these are dependent on its installation location. An external calibration target is typically used to determine the extrinsic parameters. For the mathematical relationships and the transformations using the intrinsic and extrinsic parameters reference is made at this point to further technical literature [2].

Dynamic calibration
As a result of the vehicle's proper motion the field of vision of a camera installed in the vehicle changes without the captured scene also inevitably changing. Dynamic calibration is required to record the resulting dynamically changing vehicle parameters and reproducing their effects on the transformation of world coordinates into image coordinates. The determined dynamic vehicle parameters are then passed on to the following algorithms. The dynamic parameters included on the one hand the longitudinal and lateral motion component of the vehicle, i.e., typically the inherent speed and the current steering angle, and on the other hand the rotary motions and the associated motion angles about the three vehicle axes:
– pitching about the transverse axis and the associated pitch angle,
– rolling about the longitudinal axis and the associated roll angle, and
– yawing about the vertical axis and the associated yaw angle.

The required parameters for the inherent speed, the steering angle, and the yaw angle are as a rule made available via the vehicle bus (e.g., CAN bus). The pitch and roll angles can be determined with the aid of algorithms for image-based proper-motion estimation. Image-based determination of the dynamic vehicle parameters is also called visual odometry.

Rectification

As already mentioned with calibration, an idealized camera model is used for the mathematical description of the relationship between a real point in the world and the associated pixel. In this model the pair of points (real point and pixel) and the optical center lie on the same optical beam. Furthermore, straight lines in the real world are also reproduced without distortion as straight lines.

With real cameras the assumption of distortion-free reproduction is typically not fulfilled, i.e., a straight line in the real world is reproduced crooked. The larger the aperture angle or the field of view of the camera, the greater the distortion of the optical reproduction by the objective and the more crooked real straight lines will be reproduced. Algorithms for rectifying the image are used to correct the distortion of an objective. Thanks to the use of rectification straight lines are again reproduced as straight lines and the assumption of the idealized camera model is valid again.

Radially symmetric distortions

Numerous models of varying complexity are known from the literature to describe the distortion of a real objective [2]. The simplest form of mathematical description in the form of a polynomial is radially symmetric distortion. This manifests, depending on the objective design, as pincushion or barrel distortion. With pincushion distortion the enlargement of the optical reproduction increases towards the image edge, with barrel distortion the enlargement decreases towards the image edge (Figure 12).

A frequently used mathematical representation of the radially symmetric distortion $\Delta r'_{rad}$ as a function of distance r' to the optical axis is:

$$\Delta r'_{rad} = A_1 r'^3 + A_2 r'^5 + A_3 r'^7$$

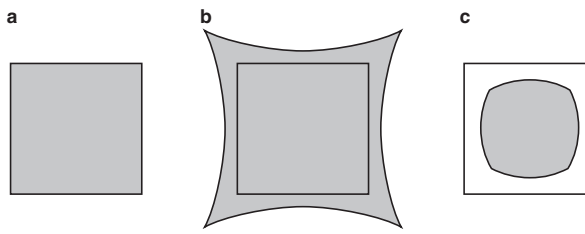
The parameters A_1, A_2, and A_3 determine the type of distortion: pincushion or barrel. Because the distortion is always specific for a particular objective, the parameters as a rule derive from an exact measurement of the objective after its production.

Further distortions

As well as radially symmetric distortion there are further distortion models, such as for example asymmetric and tangential distortion, which is caused as a matter of priority by a the lenses inside the object being brought out of center. There is no detailed description of the associated complex mathematical relationships at this point and reference is made to further literature [2].

Because a camera in the vehicle is often installed behind the windshield, the distortion caused by the window glass must also be taken into account in the rectification.

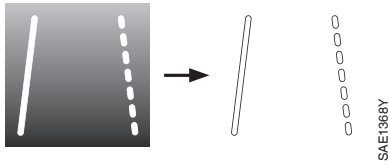
Figure 12: Radially symmetric distortion
a) No distortion,
b) Pincushion distortion,
c) Barrel distortion.



Classic image-processing algorithms
The algorithms of classic image processing are based exclusively on mathematical techniques and are divided typically into a feature extraction and a downstream object formation. The classic image-processing algorithms included in the field of driver-assistance systems typically line detection and light detection.

Line detection
For line detection the conditioned and rectified image-data stream is initially processed using a cut-off filter. This filter operation produces a gradient image from the original brightness information. In terms of signal processing a cut-off filter corresponds to a high-pass filter which emphasizes brightness jumps between neighboring pixels and suppresses brightness changes. Retained in the resulting gradient image are thus lines which exhibit at their edges a large brightness jump, such as for example a white line marking on dark gray asphalt, areas of consistent brightness are suppressed. An algorithm frequently used for edge detection is the Canny algorithm, which typically uses the Sobel operator as a filter kernel for partial derivation in the horizontal and vertical directions [1]. The Sobel operator is defined here as a 3×3 filter matrix. A typical application of the Sobel operator is shown in Figure 13. Based on the gradient image pixels with a similar gradient vector and similar intensity are brought together in line segments. After a plausibility check which contains all manner of assumptions for a valid line marking, relevant line segments are linked with valid lane markings and output as line objects. Included in the plausibility-check assumptions are for example alignment, strength and bending radius of a line as well as the parallelism of several lines. The presented algorithm facilitates the detection of all the line-based lane markings, such as for example continuous and dashed lines and line markings in different colors.

Figure 13: Application of the Sobel operator



Light detection
Light detection is as a rule used as an input variable for different light functions, such as for example automatic high-beam control. The lights of oncoming vehicles and vehicles traveling in front must be detected in a nighttime scene for this purpose. A CCL algorithm (connected-component labeling) is generally used here. This algorithm facilitates the bringing together of neighboring pixels whose brightness value is above a defined threshold value. This process is also known as clustering. In this way an image-data sequence can be segmented, i.e., pixels which belong to a vehicle light are brought together in a cluster and continued in the following as an individual characteristic (Figure 14). For the detected points of light a series of additional attributes is also established, such as for example size, center point and brightness of the respective group.

The detection of the points of light is followed by their plausibility check using different assumptions, such as for example the brightness and position of the point of light in the image. In this way relevant light objects thus detected as vehicle lights are formed and separated from irrelevant light sources, such as for example reflective road signs.

The algorithmic further development of light detection comprises, aside from the detection of front and rear headlamps, also the stop, turn-signal, tail, and hazard-warning lamps of other vehicles as well as the detection of the rotating beacons of emergency vehicles. The recognition of traffic lights also poses a further challenge for light detection.

3D reconstruction

The aim of 3D reconstruction is to restore the depth information of a captured scene because this information is lost by the camera system and the associated transition from the real three-dimensional space to the two-dimensional space of the focal plane. By restoring the depth information it is possible to create a detailed environment model in the area of the captured scene. In this way the drivable corridor ahead of the vehicle (freespace) can be determined and on this basis a precise trajectory planning can be created for the ego vehicle. A precise 3D reconstruction also facilitates – in complex scenarios, for example in roadworks areas on expressways/interstates – longitudinal and lateral vehicle guidance. A detailed environment model is absolutely necessary for autonomous driving functions. An exact calculation of the road surface is also made possible by 3D reconstruction. Irregularities such as for example potholes or speed bumps can be detected by means of this surface estimation.

Optical flow

The optical flow is an algorithmic process on the basis of which the depth information of a scene can be reconstructed.

Flow calculation

The basic principle for the optical flow is provided by a set of features which is made up of a defined number of features. A feature typically consists of two pixel sums within a defined window size, e.g., 5×5 pixels. The difference between the two pixel sums is compared with a threshold value. The result of this comparison is then in the simplest case 1 or 0, depending on whether the difference is greater or less than the threshold value. This type of feature is also called a Haar-like feature because of its similarity to Haar wavelets from signal processing. Figure 15 shows by way of example a set of features with six features in a window size of 5×5 pixels.

Figure 14: CCL algorithm

a) Real points of light,
b) Reproduction on points of light,
c) Bringing together in a cluster by the CCL algorithm (connected-component labeling).

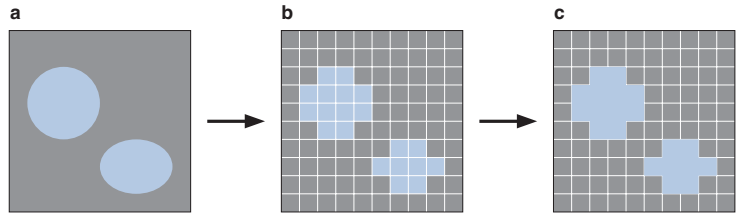
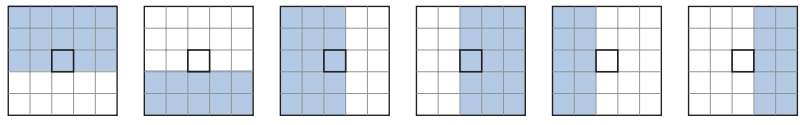


Figure 15: Set of features by way of example for 5 x 5 pixels

The examples show features for detecting horizontal edges and vertical edges.



Aside from the Haar-like features presented here there are also numerous other feature types known to science, such as for example HOG (Histogram of Oriented Gradients) or SURF (Speeded Up Robust Features). The feature size is variable and impacts directly on the computing complexity of the chosen algorithmic approach.

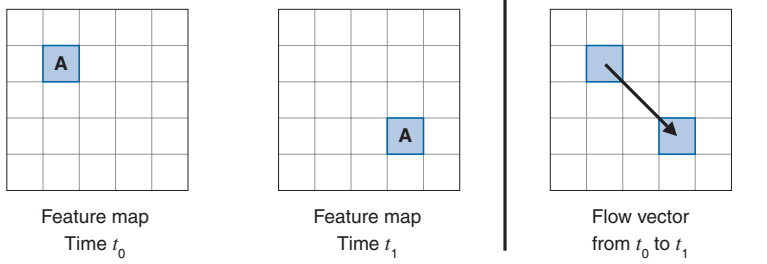
Now feature calculation is performed for each pixel within an image. Aside from the central pixel, all the pixels located in the window are included in the calculation. Based on the results of the feature calculation, each pixel is assigned a signature, e.g., 101100 for the six features from Figure 15. The aim of feature calculation is to generate as clear a signature as possible for all pixels belonging to relevant structures in the image, such as buildings, marker posts or vehicles of all types. During the feature calculation for an entire image a feature map containing all the relevant signatures is created on the basis of the original image data.

If now a feature map is calculated for two consecutive images, the pixels belonging to a structure visible in both images have the same unique signature, i.e., they correspond. This also applies if the ego vehicle moves and the position of a signature shifts from image to image. Signature-based correspondence creation, i.e., the location of identical signature values in different images, is one of the core algorithmic components of the optical flow. Starting out from the image position of a signature A at time t_1, it is thus possible, referred to the position of the same signature A at the previous time t_0, to determine a vector, which is also called the flow vector (Figure 16).

The sum total of all the flow vectors in an image is also called the flow field. In this vector field properly moving objects, such as for example a crossing pedestrian, can be clearly detected because the flow vectors on a properly moving object clearly differ from the vectors of the environment. Because the optical flow does not contain any object assumptions, i.e., the algorithm does not know what pedestrians or other properly moving objects look like, the optical flow facilitates a generic object detection.

Within the framework of scientific papers numerous algorithms for calculating the optical flow have been published in the past few years which aside from the algorithmic complexity also differ in how many flow vectors can be calculated from the image data. A distinction is essentially made here between algorithms which frequently generate very many flow vectors on the basis of interpolation and in this way can create a dense flow field and algorithms which only generate a fraction of the flow vectors and can therefore create only a sparse flow field.

Figure 16: Optical flow
A Signature.



Structure from Motion

The position of a three-dimensional world point in the two-dimensional focal plane can be calculated with the aid of static calibration. For a known pixel, however, it is conversely possible to calculate only a straight line in three-dimensional space on which this point lies. The straight line is also called the line of sight. All the world points which are on this line of sight are projected onto a common point in the focal plane. The depth information of a point in the real world is therefore lost in the reproduction in the two-dimensional space of the focal plane, i.e., based on a pixel the distance of the corresponding point in the real world cannot be determined.

To determine the depth information of a point in the real world, it is necessary to capture this point from two camera perspectives. It is thus possible to calculate two non-identical straight lines in three-dimensional space at whose crossing point ultimately the real world point is situated. Algorithms for calculating the optical flow use as explained previously two captures at different times for the creation of the flow vectors. When the ego vehicle moves, the two capture times correspond to two camera perspectives at different times.

For a static structure in the real world it is now possible to determine two non-identical lines of sight on the basis of which the real distance of the structure can be calculated. For this purpose, aside from the flow vector, the parameters for the vehicle's proper motion must be know, for example how far the vehicle has moved between the two capture times. It is thus possible to calculate depth information for all structure in the real word for which a unique signature has been allocated and which consequently can be found in the image data of both capture times.

A depth map of the captured scene is created in this way. This flow-based technique of 3D reconstruction is also called “Structure from Motion” (SfM).

A drawback of flow-based 3D reconstruction is however the dependence on the ego vehicle's proper motion, i.e., when the ego vehicle is stationary, flow-based 3D reconstruction of the captured scene is not possible.

Stereo vision

Another algorithmic technique for reconstructing the depth information of a scene is stereo vision, i.e., a scene is captured at the same time by two image sensors. This equates to human sight with two eyes. Like the optical flow, this technique too is based on signature-based correspondence creation. To this end a feature calculation is performed in the first step as with the optical flow in order to generate signatures that are as unique as possible for real structures. Because of the assigned signature a structure in the real world can then be found in a captured image. Stereo vision, however, does not use like the optical flow the signature-based correspondence creation on the basis of two consecutive images of the same camera, but on the basis of two images captured at the same time by independent cameras. If the position of a signature A is known for the optical flow at time t_1, this signature A referred to the time t_0 may have moved in any direction horizontally and vertically.

For stereo vision a simplification can be assumed under certain preconditions so that there must be a signature in both camera images within the same image line, which is then called the epipolar line. With the aid of simplification by the principles of epipolar geometry the algorithmic effort for the correspondence search can be significantly reduced. To be able to apply the simplification, it is necessary to establish and maintain the stereo normal case for a stereo camera system (Figure 17).

The stereo normal case requires firstly a parallel alignment of the focal planes of two identical cameras, i.e., cameras with

identical intrinsic parameters. In addition, the focal planes must be parallel to the base line, which connects the optical centers of the cameras. The length of the base line between the optical centers is also called the base width of a stereo system. In addition to the parallel alignment of the cameras the optical centers of both cameras must be at the same height. To completely satisfy the simplified epipolar geometry, the pixels of an image line in the left camera image must correspond to the same image line in the right camera image. Therefore a distortion-free image of the individual cameras is required as well as intrinsic and extrinsic calibration.

The function of a camera system in the vehicle must be guaranteed over a wide temperature range and is subject to constant vibration and aging effects. To this end a camera system always has production-related manufacturing tolerances and installation-related installation tolerances. To be able to guarantee compliance with the stereo normal case also for a stereo system installed in the vehicle, as a rule an algorithmic component extending beyond the mentioned intrinsic and extrinsic calibration and rectification is required to establish and maintain the stereo normal case. This component is also called stereo calibration.

A world point P captured by both cameras is now reproduced taking account of the stereo normal case in the focal planes of both cameras on the identical image line, i.e., on an epipolar line (Figure 18). The horizontal distance between the pixel and the optical center of the left camera taking account of the focal length f and the base width B of the stereo system is governed by the following relationship:

$$x_L = \frac{f}{Z} X .$$

By analogy, the horizontal distance between the pixel and the optical center of the right camera is governed by the following relationship:

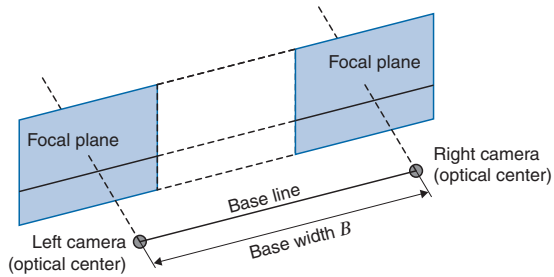
$$x_R = \frac{f}{Z} (X - B) .$$

The difference between the horizontal position of the pixel in the left camera image and the horizontal position of the pixel in the right camera image is called the disparity d and describes the relationship between the reproduction of the world point P and the depth Z:

$$d = x_L - x_R = f \frac{B}{Z} .$$

The depth Z corresponds to the distance between the world point P and the optical center of the reference camera system,

Figure 17: Stereo normal case



i.e., in this case the left camera. The distance to the world point P is thus inversely proportional to the disparity d. In this way it is possible to calculate the disparity for all the visible and clearly corresponding world points in the two images and create a disparity map which contains the depth information of the captured scene. Generally: The greater the disparity the closer an object is and the smaller the disparity the further away an object is.

Comparison of optical flow and stereo vision

Stereo vision has a number of advantages over 3D reconstruction based on the optical flow, but also disadvantages for an application in the automotive sector. The stereo-based measurement method is constantly available compared with the optical flow and is not dependent on the vehicle's proper motion. The depth calculation of the flow-based method on the other hand requires as accurate an estimation of proper motion as possible and knowledge of the capture times of the individual images because these have a direct bearing on the accuracy of the depth estimation.

Stereo vision offers a direct geometric possibility of depth calculation and therefore as a rule delivers greater measuring accuracy.

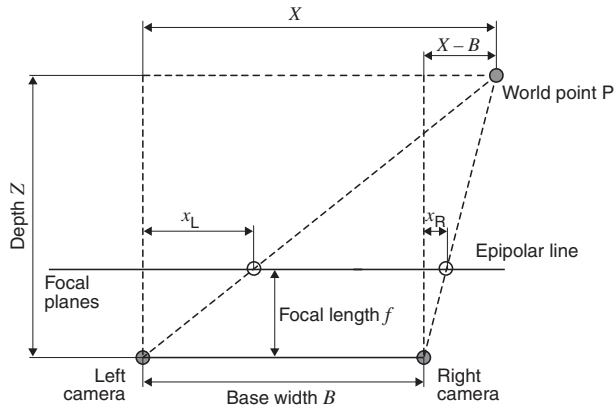
A drawback of stereo vision is the need for a second camera, which typically entails much higher costs for a stereo camera system. The optical flow offers analogously to stereo vision the possibility of calculating depth information on the basis of the vehicle's proper motion with a generally cheaper mono camera system. For this reason the flow-based 3D reconstruction technique “Structure from Motion” is frequently also called “Motion Stereo” [2].

Machine learning techniques

In the field of computer vision a typical task for an algorithm is to search for specific objects in the image, such as for example pedestrians, bike riders, motorcycles, cars, and trucks. Detecting these objects is a challenging task for a computer system, which is typically solved with techniques from the field of machine learning, which is a sub-area of artificial intelligence.

Machine learning techniques enable a computer system to process specific tasks without having been explicitly programmed for these beforehand. The algorithms learn on the basis of image-data sequences how they can best manage

Figure 18: Stereo disparity



a specific task. These image-data sequences involve so-called training data in which the objects to be detected were identified beforehand. The identification of objects is also known as labeling and is the basis for supervised machine learning. Based on the labeled training data the algorithm generates a mathematical model that enables its to detect relevant objects within an image with a high degree of probability. Not only does the algorithm detect these specific objects in the known training data, but it is also in a position to detect these objects in new and previously unknown image data. The greater the labeled training-data volume the more probable it is that the algorithm will detect a specific object hitherto unknown to it. What is important here is not only the amount of training data but also as high a variance as possible within the training sequences, such as for example the coverage of all manner of light and weather conditions.

Because the training data are typically stored on servers or a data cloud, the volume of training data is an important cost factor. Ideally, a computer system would detect and classify 100 % of all the relevant objects within a scene. However, this is not possible in the real world on account of their complexity. Thus, the aim of object detection is as high a detection rate as possible with at the same time a low error rate. With machine learning techniques the achievable detection rate is decisively dependent on the training-data volume used and thus the associated data costs.

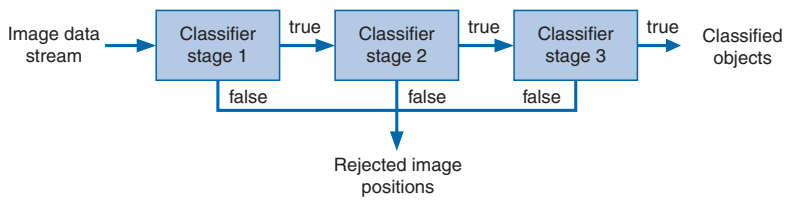
Cascading object classifier

An example of the implementation of a machine learning technique is the cascading object classifier, which is typically based on a supervised learning technique. Here for each item of input data the labeled training-data volume is used to defined the result to which the classification must lead. Typically a cascading object classifier is used to detect and object class such as for example pedestrians, bike riders or motorcycle riders.

A cascading object classifier generally consists of a number of classifiers which are linked in the form of a decision cascade. This is then applied to each pixel of the incoming image-data stream. If the decision cascade has been fully run through, there is a very high probability that there is a relevant object at this image position. If the cascade has not been fully run through, there is a very high probability that there is no relevant object at the processed image position. The decision cascade is then canceled early and the current image positions are rejected. Through the early rejection of non-relevant positions the cascading structure facilitates an optimization of the total running time of the object classifier because all the existing classifier stages doe not have to be calculated for each image position (Figure 19).

Each classifier stage consists of a decision tree which in turn is made up of individual linked features. The feature type used involves in the simplest case, as in the implementation of the optical

Figure 19: Cascading object classifier



flow, Haar-like features. The manner of the links within the decision tree and the concrete elaboration of the features is determined by the previously specified learning technique (Figure 20). Numerous algorithmic techniques are known in the literature for implementing a cascading object classifier [3].

Deep learning

In the area of machine learning the basic relationship that a larger volume of labeled training data leads to a higher detection rate of the algorithms used applies. In particular, this applies to the deep-learning techniques which, when compared with classic machine learning techniques such as for example a cascading object classifier for an increasing volume of training data, can achieve a higher detection performance (Figure 21).

The deep-learning techniques are based on artificial neural networks the structure of which on an abstract level are based on the information processing in the human brain. The algorithms uses here a number of processing layers which process the incoming raw image data in stages. The deep-learning algorithms learn independently which calculation parameters must be used on which level to achieve an optimal detection result.

One implementation possibility for an artificial neural network is the deep neural network (DNN), in which data processing is performed in only one direction from input layer to output layer (Figure 22). The DNN thus works from a signal-theory standpoint exclusively on the forward path (feedforward). For the area of computer vision convolutional neural networks (CNNs), which are a specific form of deep neural network, are typically used. In the internal processing layers of a CNN discrete convolution operations are used which when compared with the general matrix multiplication of a DNN can be carried out with simpler filter operations and can thus be implemented with less effort in an embedded system. The processing layers between input and output layers are also called hidden layers.

Figure 22 shows a strongly simplified representation of a typical CNN architecture from the field of computer vision. The CNN network architecture depicted is based here on the previously rectified image data which are called in this connection the input layer. The convolutional layer is the central unit of the CNN architecture and contains an adaptable filter which is applied using discrete convolution to the input image data. In this way a specific activation threshold is determined

Figure 20: Classifier stage with Haar-like features

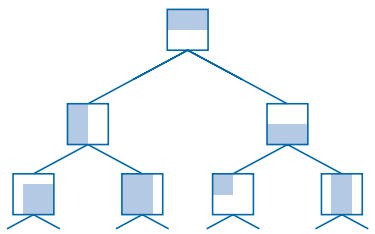
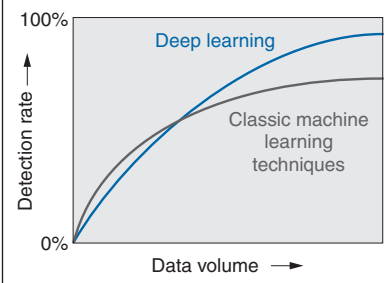


Figure 21: Relationship between detection rate and data volume



for each filter kernel corresponding to an artificial neuron. Depending on the CNN result the values of the filter kernels are independently optimized by the network, i.e., the neural network learns independently. In the pooling layer the existing data volume is decreased by combination of the filter responses and the size of the filter kernels is reduced. In the ensuing fully-connected layer all the neurons of the pooling layer are connected with all the neurons of the following output layer. In a CNN-based classifier the number of neurons in the output layer as a rule corresponds to the number of object classes to be classified.

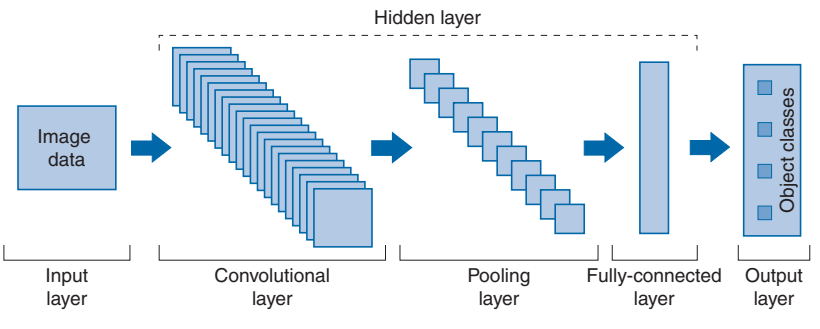
There are as a rule in a CNN a number of interlinked units consisting of convolutional and pooling layers which thus form a multi-stage or else deep neural network. For the implementation of a CNN in a driver-assistance system the CNN network architecture is typically adapted to the architecture of a suitable hardware accelerator within the embedded system.

In the field of driver-assistance systems a CNN is as a rule taught in during the development process and adopted in series in a defined state. Training takes place under monitored condition, i.e., the object classes for the output layer of the neural network are specified during the learning phase. The validation of an algorithm based on deep learning is the object of the validation. This is based as a rule on a volume of validation data that are not dependent on the training data.

A CNN does not require any additional conditioning of the input image data when compared with other algorithmic techniques such as for example line detection based on cut-off filters. In the classic image-processing techniques the filter kernels must as a rule be developed on an application-specific basis; a neural network, however, learns these filters implicitly.

One of the typical applications for a CNN is semantic segmentation, which assigns one of the learned object classes to each pixel. In addition, a CNN facilitates the detection and classification of objects such as for example road signs, traffic lights, and lane markings and roadsides of all kinds (any boundaries). Single-task networks, which process only one of the tasks mentioned, and multi-task networks, which facilitate the performance of a number of tasks within a CNN, are possible here. The architecture of single-task and multi-task networks is likewise the object of current research in the field of computer vision.

Figure 22: Strongly simplified representation of a typical CNN architecture



Data interface
Objects and measured data are typically transmitted via the computer-vision data interface. The objects detected and classified by the classic image-processing algorithms and the machine learning techniques include for example lines, points of light, vehicles, and pedestrians.

In the case of measured data the depth map determined with the aid of 3D reconstruction is transmitted for example.

References
[1] W. Burger, M.J. Burge: Digitale Bildverarbeitung. 3rd Edition, Springer, 2015.
[2] R.I. Hartley, A. Zisserman: Multiple View Geometry in Computer Vision. 2nd Edition, University Press, Cambridge, 2004.
[3] H. Süße, E. Rodner: Bildverarbeitung und Objekterkennung: Computer Vision in Industrie und Medizin. Verlag Springer Vieweg, 2014.

Sensor-data fusion

Introduction

Perception of the environment
People use their sense of sight when driving a car to discern the information pertaining to environment detection. Thus, the conclusion could be made that an autonomously driving vehicle only needs one video sensor to perceive the environment (or surroundings). Current autonomously driving platforms, however, have a multitude of sensors which operate according to different measurement principles. Perception of the environment is then conducted on the basis of the "fused measurements" of these sensors. Sensor-data fusion is concerned with the question as to how measurements from a variety of sensors should be combined to enable a scene to be optimally perceived. Sensor data thus refer to the acquired measured data and not the values from data sheets (e.g., measuring range, accuracy).

Sensor-data fusion in the automated-driving environment
Sensor-data fusion is essential in the field of automated driving for a number of reasons. Humans have as part of their evolution developed highly effective and efficient perception based on visual stimuli. We are currently not yet at the stage where we can emulate this so well and computation-efficiently that it can be implemented in a vehicle with limited energy resources. Current video sensors are stretched to their limits particularly in adverse-weather conditions, for example due to the low signal-to-noise ratio (SNR) and thereby limited detection capability.

Furthermore, an automated-driving solution must be better than a human and at the same time safe in order to be accepted by society. The processing overhead in perception can be reduced by having the scene measured very well and in such a way that the separability of interesting from uninteresting scene content is consistently as big as possible. This happens currently by using numerous sensors to measure the environment. In addition, fused measurements of numerous sensors also increase safety, above all in the event of a sensor failure.

Glossary
Special terms are used in the field of sensor-data fusion and will be explained here.

Sensor space
The sensor space contains the depiction of the environment based on the sensor measurements, e.g., pixels or images for a video sensor and point clouds for a lidar sensor.

Feature vector
This is a vector with features. With regard to a pattern a feature is a scalar by which a particular aspect of the pattern is described.

Inference
Inference denotes the acquisition of information by reaching a conclusion. Within the framework of a stochastic consideration it is concerned with the calculation of joint and conditional probabilities, e.g., in the case of a classifier the conditional probability of the class output is calculated depending on the feature-vector input.

Perception
Perception refers to a subconscious process of individual information and perception processing which gives rise in the consciousness of the information recipient to images of perceived partial aspects of reality.

Heuristics
Heuristics is the art of obtaining probable statements with only limited knowledge.

Sensors

There are currently in the automated-driving environment a number of types of sensors whose measured data are fused to perceive the scene. Each sensor measures certain specific properties of objects (e.g., of road signs, of a vehicle parked at the roadside, of pedestrians). Using these properties it is then possible to separate the objects from the background and differentiate them from each other – on the assumption that these differences are found in the sensor measurements.

Sensor types

Video sensors

Video sensors (image sensors, imagers) operate with electromagnetic waves in the visible-light range. A video sensor is able to convert photons to charge or to voltage. It generates a signal depending on the intensity and color composition of the light that is incident upon the sensor.

There are two types of objects: light sources and reflectors. In both case the composition of the light that comes from an object is dependent on specific object properties (e.g., surface structure).

Lidar sensors

Lidar sensors operate with infrared light. Unlike a video sensor, lidar is an active sensor. Lidar sends light beams with a certain distribution in the scene and receives the reflections back. The lidar uses the delay with which the emitted light after being reflected by the object is received by the sensor to measure the distance to the objects in the scene. A 3D representation of the scene is thus produced. The objects are then distinguished via their geometric structure. Modern lidar sensors are also able to measure the speed of objects.

Radar sensors

Radar sensors operate with electromagnetic waves in the radio range. Radar is likewise an active sensor, emitting a signal and measuring the reflection of the signal by objects in the scene. Essentially the distance to the object is measured by way of the delay.

Radar sensors also measure the speed of the object and properties of the reflecting object surface. Compared to lidar, radar typically delivers a sparser 2D range map and not a dense 3D representation of the scene, but the quality of the radar measurement is significantly less susceptible to weather conditions. Different technical approaches can be used to increase the resolution of radar sensors significantly.

Ultrasonic sensors

Ultrasonic sensors operates with sound waves above the audible frequency range. They are also active sensors. Ultrasonic sensors have a very short range compared to video, lidar, and radar, but are low-priced and robust against weather conditions.

Perception

To perceive a scene, the information from the environment, just as it is measured by a sensor, is analyzed in a first step and relevant scene content (e.g., objects) is detected and tracked over time. In a second step the objects are placed in relation to each other. This second step generates additional information, which, together with the information extracted in the first step, describes the scene in full.

En route from the sensor measurement to the detected object the scene information passes through various processing steps (Figure 1). To improve object detection and tracking, information from a number of sensors is combined at different points along this processing chain.

Thus, sensor-data fusion is concerned with the processing of information which is obtained from the measurements of different sensors. The aim is to separate important scene content robustly and with high quality from unimportant content and to track the important content over time.

Representation of the driving scene

Currently most algorithmic solutions of the fusion tasks follow two architectures, which differ in the way in which information from the driving scene is depicted.

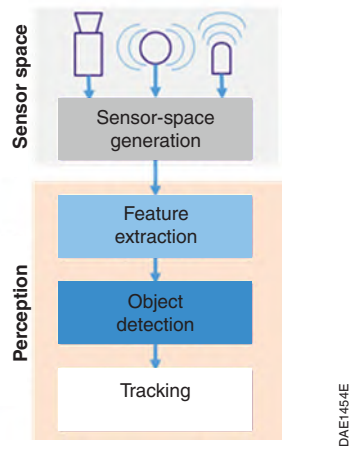
Grid approaches

Grid approaches are based on discretizing the driving scene into typically non-overlapping areas. At each point in

time a decision is made on the basis of the sensor measurements as to whether one of these areas is filled by an object. Individual areas can then be aggregated to objects in further steps [1].

Object-model approaches
Object-model approaches for their part use from the outset an explicit representation of the objects. This representation is of a geometric nature and is usually based on the principle of the enclosing polytope. It contains object properties such as size, orientation, position, and dynamics and is recalculated at each point in time on the basis of the sensor measurements.

Figure 1: Functional block diagram of early fusion



Fusion types

Depending on the position along the processing chain at which the information from the sensors is combined, a distinction is made between three types of sensor-data fusion.
– Early fusion: Here sensor measurements are directly combined with each other. This happens directly at the sensor output, prior to any further processing.
– Mid fusion: Here certain perception-related information-processing steps are carried out sensor-specifically first before the results of these steps are combined.
– Late fusion: Here objects are perceived sensor-specifically before they are combined.

Early fusion
En route from the sensor measurements to a detected object the scene information passes through a number of processing steps. In this way measurements from the sensor space are aggregated to feature vectors before it is decided whether and what kind of an object is situated at a certain point.

In the case of early fusion the sensor space extends across all the available sensors. Subject to this common sensor space is a model which describes how individual sensor measurements are to be combined with each other. The parameters of the model can be set on the basis of a-priori knowledge or learned from data. The block diagram of this type of early fusion is shown in Figure 1.

Traditional approaches
Traditional approaches to early fusion contain a separate step in which the sensor space is explicitly generated and the model parameters are set with reference to different considerations. This step is the basic principle on which the other information-processing steps function.

Learning method
If there is to be an underlying learning method, there is the possibility of combining the sensor-space generation with subsequent and likewise learning-based information-processing steps. Here the parameters of the respective algorithmic blocks are learned together. In this case the sensor space is only implicitly generated. Better results can be obtained by optimal mutual adaptation of the steps that are based on one another. However, it is in this case more difficult to follow which type of information is used where and what effect each step has along the processing chain. Neural networks offer the opportunity to implement such a procedure relatively easily.

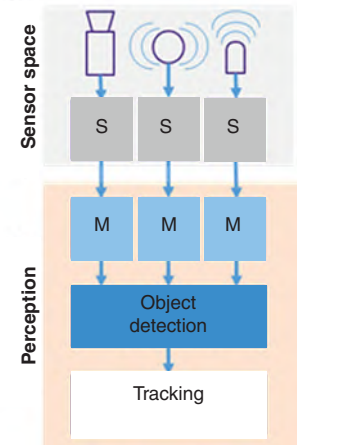
Discussion
Perception functions like an information funnel. This extracts from the superset of scene information which the sensors measure only the information relevant to the automated-driving tasks. There information is filtered out at different points by different mechanisms. One of these mechanisms provides that potential object information with a reduced SNR is ignored. The upshot of this is that incorrect detections but at the same time also some correct detections are prevented.

Early fusion has the potential to achieve optimal perception performance. This is primarily because complementary information from all sensors is used without discarding object information with a reduced SNR from one sensor. Even if an object exhibits information with a low SNR for all sensors, there is in the case of early sensor-data fusion a good opportunity to detect the object by analyzing the dependencies between the measurements in different sensor spaces.

Early fusion also has certain drawbacks. For optimal results the sensors must be very precisely synchronized. This is often difficult because the times of measurement can differ very markedly from sensor to sensor. The calibration requirements are also very stringent in the case of early fusion and the hardware requirements increase because larger volumes of data have to be processed simultaneously. Calibration is the name of the process of determining sensor parameters. Furthermore, already the failure of a single sensor can lead to problems in the subsequent processing chain.

Another difficulty of a practical nature is that the sensor interface is not yet standardized. The ISO 23150 standard TC22/SC31 [8], which is intended to standardize this interface in future, is currently in the preparation stage. Not all the sensors make a raw-data interface available or meet the necessary synchronization and calibration requirements.

Figure 2: Functional block diagram of mid fusion
S Sensor-space generation,
M Feature extraction.



Mid fusion

In the case of mid fusion the sensor spaces of the individual sensors are examined for conspicuous regions and if necessary features are extracted before these are combined for object detection. Similarly to early fusion objects are in this way detected on the basis of an overall-sensor description of the scene; however, the search for conspicuous regions is sensor-specific. Figure 2 shows a typical block diagram for this type of fusion.

Discussion

Mid fusion has the potential for improved performance with similar, but less pronounced drawbacks of early fusion. Thus the synchronization, availability, and hardware requirements are less stringent because not all the measurements of the various sensors have to be processed simultaneously. However, the difficulties relating to the interface also remain for mid fusion.

Late fusion

In the case of late fusion objects are first detected sensor-specifically before they are aggregated on an overall-sensor basis. One option of aggregation entails tracking the sensor-specifically detected objects jointly with the aid of a specially adapted measurement model over time. Figure 3 shows the block diagram of this type of fusion.

There are different heuristics to carry out the aggregation. Such approaches also work when the objects in the scene are sensor-specifically detected and tracked.

Discussion

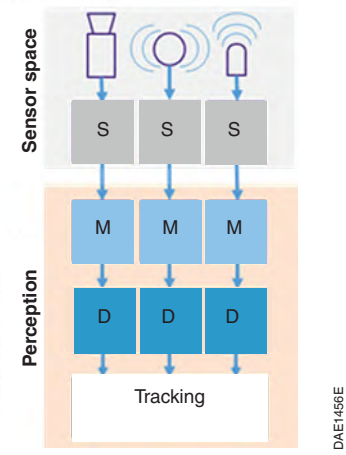
Later fusion cannot achieve the performance of early fusion because information which would only be detected in the sensor network is discarded here. Late fusion cannot achieve the quality of mid fusion either because typically sensor-specific information in the low SNR range is ignored in order to reduce the number of incorrect detections. However, late fusion does not exhibit the drawbacks of early and mid fusion. Because it is objects and not sensor measurements or scene descriptions that are processed, the data rate is very low. The synchronization and calibration requirements are also low and sensor availability is non-critical because in the event of a sensor from the fused network failing the scene can nevertheless be perceived by having the objects of the other sensors analyzed.

Application of the fusion types

Late fusion is currently the most frequently used solution in the field of driver-assistance systems. The availability of greater computing capacities and the need for higher performance in the field

Figure 3: Functional block diagram of late fusion

S Sensor-space generation,
M Feature extraction,
D Object detection.



of automated driving means that the focus of development is on primarily mid fusion. Early fusion presents itself above all as the benchmark and for task with the highest performance requirements.

Information basis of object detection
Depending on how information is introduced over time in the decision problem of object detection, there are a number of ways of fashioning this step algorithmically. Deep neural networks with special architectures [2] play a particular role here.

Sensor-specific measurements (or features generated from them) can be jointly tracked across conspicuous regions. There is thus a history of sensor-specific information available to object detection. Alternatively, only information which was collected across overall-sensor conspicuous regions from sensor-specific measurements without a history is used. Particularly for early and mid fusion the decision about the procedure for object detection leads accordingly to an adaptation of the functional block diagram. The aggregation step can then solve different tasks, such as for example tracking or the consolidation of tracking results.

AI-based sensor-data fusion

Artificial intelligence
Artificial intelligence (AI) is concerned with the question as to how a machine has to be programmed or designed in order to complete tasks for which a human must use intelligence. Intelligence is defined as the capacity to acquire information (e.g., by perception or by inference), to store and use this information as knowledge, to handle this optimally, and to adapt to changes by learning. In general terms, the aim here is to support or even replace humans through the use of AI. Thus, AI is the cornerstone of autonomous driving.

Sensor-data fusion is important in order to detect and track objects within the framework of perception, where detection refers both to the division into important and unimportant objects (i.e., detection) and to the specification of the object type (i.e., classification). Different sensors can be used to measure the scene optimally and thereby establish the basis for optimal detection and tracking.

The tracking of objects and their detection are implemented as part of perception by AI algorithms. Usually the parameters of the algorithms that implement the tracking are defined on the basis of a-priori knowledge while the parameters of the algorithms that implement the detection are learned from training data. But there are also algorithmic solutions to the tracking problem which are based on learning approaches.

The algorithmic paradigms which make it possible to carry out detection and tracking and then in each case the most common approaches in the automated-driving context will be discussed in the following. Fusion approaches differ in how the combine and implement these algorithms.

Object detection

A decision problem is solved in order to detect objects. All the sensor measurements are analyzed and a decision is made whether these originate from an object. In practice the definition of an object is highly variable. Thus, an object can be understood to be part of a set, such as for example everything that moves, or standalone, e.g., a motorcycle. The term used here is detection when the important is separated from the unimportant, where what is important is problem-dependent.

Classifiers

Decision problems are solved by classification algorithms. These algorithms implement classifiers. These are mathematical functions which constitute a relationship between an input and a discrete output. The input is a mathematical object (usually a vector) which contains a representation of the object, about which a decision is to be made. The output is connected with a label for a class to which this object belongs. In the case of object detection the sensor measurements (or quantities derived directly from them) constitute the input of the classification algorithms and the set of object types the output.

The mathematical expression of a classifier is an important design parameter. Of particular importance in view of their performance are neural networks (NN), where different types of classifiers exist. Neural networks are classifiers which are constructed by the combination of a number of simpler functions, i.e., neurons.

The parameters of a classifier can either be determined manually on the basis of a-priori knowledge or determined automatically. Automatic determination of the parameters happens as training as part of a machine learning process. The calculation of the output of a classifier for a given input is called inference.

Either data or rewards underlie the training. If exclusively input-output pairings (i.e., labeled data) are to be used during the training, the training is supervised. If the training is to use exclusively input data (i.e., unlabeled data), the training is not supervised. There is also the possibility of using a mixture of labeled and unlabeled data.

Linear classifiers and the perceptron

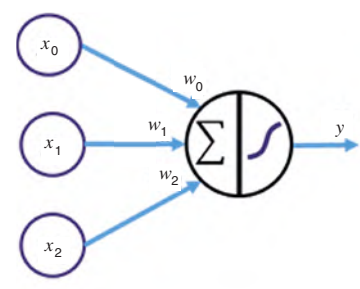
The simplest decision problem that a classifier could solve is binary. In this case the input of the classifier is to be assigned to one of only two possible classes. Because multiple-class decision can be made as a combination of a number of binary decisions, it is sufficient for the moment to consider the binary classifier.

The simplest binary classifier is based on a linear function $C : \mathbb{R}^N \rightarrow \mathbb{R}$, i.e., a first-order polynomial. This function assigns to each input vector $\vec{x}$ an output y as $y = \vec{w}^T\vec{x}$ where $\vec{w}$ is a vector with parameters.

A classification algorithm assigns the input to a class label. Because the function C supplies a real value at the output, within the framework of the classification algorithm the output y is to be mapped on a label set, e.g., {0,1}. This is done as a rule using a nonlinear function H. A linear binary classifier K thus consists of a composition of two functions $K = C \cdot H$, where C is a linear and H a nonlinear function. Figure 4 contains a graphical representation of a classifier.

The zero-level set of a linear classifier is a hyper level. Accordingly only such separating levels between the classes can be modeled in the input-vector space, which constitutes a significant limitation of linear classifiers in practice.

Figure 4: Graphical representation of a linear classifier



The perceptron is a linear binary classifier which uses the Heaviside function as nonlinearity. The Heaviside function is defined as:

$$H: \mathbb{R} \rightarrow \{0,1\}$$

$$H(y) = \begin{cases} 0, & y < 0 \\ 1, & y \geq 0 \end{cases}$$

Thus the input vector belongs to the class "1" when $y \geq 0$, otherwise it belongs to the class "0".

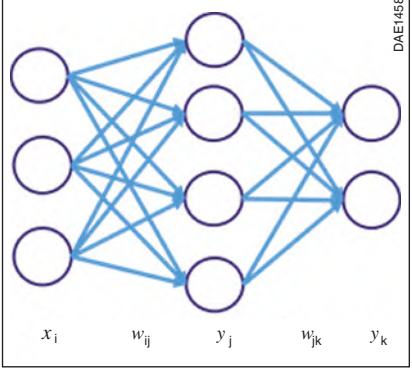
Deep neural networks

The perceptron originated from research conducted in the field of neuroscience as a model of an artificial neuron. In this context it is intuitive to combine neurons on the example of the brain in networks. Such networks can also model nonlinear interfaces and no matter which, however in practice only if they consist of neurons, where the nonlinearity in contrast to the Heaviside function is derivable like for example the sigmoid function:

$$S(y) = 1/(1+e^{-x})$$

Figure 5: Simple neural network consisting of three layers

x_i Neurons of the input layer,
w_{ij} Weights of the hidden layer,
y_j Neurons of the hidden layer,
w_{jk} Weights of the output layer,
y_k Neurons of the output layer.



Neural networks can be constructed by combining neurons in layers and building the layers on top of each other. The more layers are built on top of each other, the deeper a neural network will be. The simplest network consists of three layers:

– an input layer, whose task it is to make available only the input data of the network,
– a hidden layer, and
– an output layer, which consists entirely of neurons.

A simple neural network is shown in Figure 5. Thus the inputs x_i in the network are processed by the hidden layer:

$$y_j = \sum w_{ij} x_i ,$$

while the output layer y_k processes the outputs y_j of the hidden layer:

$$y_k = \sum w_{jk} y_j .$$

Such flat neural networks can already approximate all the possible nonlinear functions and are in this sense universal approximators. In practice, however, it has emerged that deeper neural networks process the information more efficiently than flat neural networks because they achieve comparable results with few parameters.

Compared to a flat neural network a deep neural network can aggregate information from the input at different abstraction levels and thereby eliminate interference factors, which have a detrimental influence on the decision quality. In the end the aggregation leads to the calculation of ever better features of the input data to the effect that these features become important supports for a good decision as to the class of an object. From this perspective a deep neural network consists of a feature-extraction transformation whose output is incorporated in a classifier, i.e., the output of the transformation is used as the input of a classifier.

Features which are derived from the input data by different calculations can simplify the classification problem. For example, motorcycles are to be distinguished from passenger cars on the basis of images. An important feature could

be the number of wheels. Thus, instead of analyzing pixels, the number of wheels is considered in order to decide on the class of the input object. However, the input must be transformed in order to reach the features that are of interest. The major practical advantage of deep neural networks is that this transformation is also learned during the training with the classifier parameters.

Supervised training

The case in which the training occurs on the basis of a set of labeled data is considered in the following because this method is currently through its performance the most widely used in the automated-driving environment and plays a particular role in perception.

The training is implemented as an optimization method. The parameters of the classifier are calculated as the argument of a function, which implements a specific measure, for which this function achieves its optimum. There are different measures which can be used. These measures are defined within the framework of learning theories.

Ideally a learning theory should result in a measure that minimizes the probability of an error relating to inference time. Because this is difficult or even impossible to determine, the measures usually consider the complexity of the classifier and the errors that it makes on the existing data record. The underlying intuition is that fewer complex classifiers generalize better. Where a classifier that generalizes wheel decides for new input data similarly well as for the data on which it was trained.

For deep neural networks an attempt is made in the case of preset complexity to minimize the error on the training data record. Because well designed deep neural networks are universal approximators, minimizing the error on the training record always works. However, the neural network then memorizes and cannot generalize. This contradicts the previous generalization assumption that the performance on the training data is reflected on previously unseen data. The result of this is that usual learning theories simply do not correctly grasp the generalization capacity of neural networks, which in turn compels the application of different heuristics. In order to that effect to maintain the generalization capacity, the quality on an additional data record, the validation data record, is observed during the training. The parameters of the neural network are set with the aid of the training data, but the training is interrupted as soon as the error on the validation data rises. There are also other different heuristics which improve the generalization capacity particularly for deep neural networks.

During the trainings an optimization problem is solved with the aid of a gradient method. Derivations are to be determined for this purpose. That is the reason why perceptrons, which use a non-derivable nonlinearity, are not suitable for neural networks.

For neural networks the training is conducted efficiently with the aid of the “backpropagation algorithm”. For deep neural networks special measures are to be taken to guarantee the successful convergence of this algorithm.

Tracking

The sensors measure the scene in predefined time intervals. A measurement is available in each case at the end of a time interval. Through tracking an object is to be found again in each measurement and the properties of an object (e.g., its speed and orientation) observed over time. The tracking problem is aggravated by the fact that the movement of the object in the time interval between two measurements is unknown. In addition, sensor noise prevents a simple assignment of a measurement to an object.

The tracking problem is solved by the analysis of a chronological sequence of measurements. There are different approaches to conducting this analysis. Firstly, the modeling of the chronological sequence is to be considered as a discrete stochastic signal. The parameters of this model are established on the basis of a-priori knowledge.

What is notable in this connection is that certain object properties cannot be observed directly. This happens indirectly by way of an estimation of the attributes from the measured values over time.

Stochastic signal

A stochastic signal $\xi_t^{(k)}$ (Figure 6) depends on two parameters: the realization index k and the time index t. For a certain time index t_1 $\xi_{t1}^{(k)}$ is a random variable. A realization i corresponds to a time signals $\xi_t^{(i)}$.

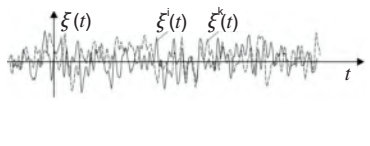
A random signal can be described by the joint probability $p(x_1, x_2,\ldots, x_n)$, where $x_1 = \xi_{t1}^{(k)}$. For tracking applications in the field of automated driving x_1 is usually a vector (i.e., the state vector), which aside from the position of the tracked object at time t_1 can also contain the speed and other properties of the object.

To solve the tracking problem, it is necessary to calculate the state vector x_t at time t. For this purpose first the probability-density function $p(x_t)$ is determined and then x_t as the argument, for which $p(x_t)$ reaches the maximum. Assuming a Gaussian distribution x_t is calculated using the expectation operator.

The tracked object moves and does not change arbitrarily over time, its state vector at time t depends on the state vector at time t–1. Furthermore, the full history of the object is known and thus its current state can be better extrapolated. The probability of x_t can be calculated as a function of the full history of the object, i.e., what is sought is the conditional density $p(x_t | x_{t-1},\ldots, x_1)$.

This conditional density can be derived from the stochastic model $p(x_1,\ldots, x_n)$ of the tracking problem, i.e., inference is carried out in the model. The basis of inference is the property that a joint probability can be shown as the product of a number of factors, where each factor is either a conditional probability or a different joint probability.

Figure 6: A random signal – two realizations i and k are represented



Stochastic-graphical models

By showing a stochastic model in graphical form, it is possible to make an inference in an intuitive way. One method for this is to use so called stochastic-graphical models [3], also known as probabilistic graphical models (PGMs).

A graphical model reflects the way in which a joint probability can be represented as the product of conditional probabilities, together with the dependencies between different components (i.e., random variables) of the model. Each node represents a random variable or a set of random variables and each connection a dependency.

To solve the tracking problem it is assumed that the current state vector only depends on the previous one and only via it on all other previous states before last. The corresponding graphical model is shown in Figure 7.

Kalman filter

To solve the tracking problem it is not possible to observe the state vectors directly. Instead, each state is measured by a sensor and these measurements which are generated from the states are observed. Because the sensor measurements are subject to noise and the noise is not deterministic, each measurement is itself modeled as a random variable y_t. The complete graphical model of the tracking problem is shown in Figure 8.

The tracking problem is now formulated as follows: Determine the state x_t from the previous state x_{t-1} (i.e., if the previous state is known) and all the observed measurements $\{y_1,\ldots, y_t\}$.

The algorithmic solution to this problem is the Kalman filter [3]. The Kalman filter operates on a number of simplifying assumptions. First, the assumption is made that all the random variables of the model in Figure 8 are subject to a Gaussian distribution. In this way the inference can be calculated in closed form. It is further assumed that the state process is a linear dynamic system: $x_t = Fx_{t-1} + v$, where v is Gaussian noise $N(0, V)$ with mean value 0 and variance V. Finally, it is assumed that the measurement model which connects the observations y_t with the states is also linear: $y_t = Cx_t + w$, where $w \sim N(0, V)$. The

parameters F, V, C and W can also be time-dependent.

On these assumptions x_t is estimated as a mean value of $p(x_t|y_1,\ldots,y_t)$, i.e., on the Gaussian assumption of the most probable realization of the random variable x_t. Thus x_t is estimated by the following relationship:

$$m_t = \int x_t \, p(x_t \,|y_1,\ldots,y_t) dx_t \,.$$

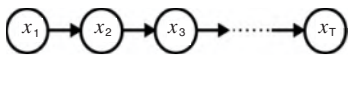
Tracking in practice

The Kalman filter as a rule delivers an acceptable solution to the tracking problem. In many cases, however, it is necessary to dispense with certain assumptions. If for example Gaussian assumption has to be dispensed with, the solution to the tracking problem must be provided by the particle filter or by variation approaches [3]. If the assumption of linearity falls away, the tracking problem can be solved by the extended Kalman filter or the unscented Kalman filter [3].

In practice it is often necessary to confronted the fact that the parameters F, C, V and W are time-dependent. The variances V and W should then be estimated from the scene via alternative paths from the available evidence. A similar produced is also followed to determine the measurement model.

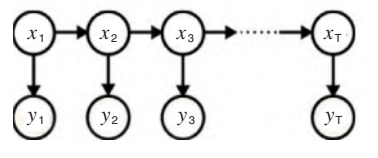
The parameters of the state model F are themselves modeled as a random variable if they are time-dependent. The graphical model is adapted accordingly. In this case the inference is however even more difficult. Simplifications exist in the form of multi-model approaches. These conduct a discretization of the functional space of the state model. The most well known of these approaches is the Interacting Multiple Model (IMM) [4].

Figure 7: Stochastic-graphical model for the tracking problem



Compounding the situation is the fact that in practice usually a number of objects are to be tracked simultaneously. The association of new measurements with a specific object is then complicated. This association problem has various solutions, many of which are problem-dependent. The most well-known general approaches are “Nearest Neighbour”, “Joint Probabilistic Data Association”, “Multi Hypothesis” [5], and “Random Finite Sets” [6]. Of increasing practical importance in this connection is the “multi-Bernoulli filter”, which elegantly addresses this problem against the backdrop of stochastic-graphical models [7].

Figure 8: Stochastic-graphical model of the tracking problem including modeling of the measurement process



References

[1] A. Elfes: “Using occupancy grids for mobile robot perception and navigation”, Computer, Vol. 22, No. 6, pp. 46 … 57, 1989.

[2] M. Liang, B. Yang, and Y. H. R. U. R. Chen: “Multi-Task Multi-Sensor Fusion for 3D Object Detection”, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, 2019.

[3] C. M. Bishop: Pattern Recognition and Machine Learning, Springer, 2006.

[4] E. Mazor, A. Averbuch, Y. Bar-Schalom, J. Dayan: “Interacting Multiple Model Methods in: Target Tracking: A Survey”, IEEE Transactions on Aerospace and Electronic Systems, Vol. 34, No. 1, 1998.

[5] Y. Bar-Shalom, X. R. Li: Multitarget-multisensor Tracking: Principles and Techniques, Storrs: YBS, 1995.

[6] R. P. S. Mahler: Statistical Multisource-Multitarget Information Fusion, Artech House, Inc., 2007.

[7] B.-T. Vo, V.-N. Vo, A. Catoni, “The cardinality balanced multi-target multi-Bernoulli filter and its implementations”, IEEE Transactions on Signal Processing, Vol. 57, No. 2, pp. 409…423, 2009.

[8] “International Organization for Standardization”, [Online]. Available: https://www.iso.org/committee/5383568.html.

Vehicle navigation

Navigation systems

Determination of the position of a vehicle is based primarily on using the GPS satellite positioning system (Global Positioning System). The navigation system compares the determined position with a digital map and uses this map to calculate the route to the specified destination.

Vehicles can be equipped at the factory (original equipment) with permanently installed navigation devices (Figure 1). Portable navigation devices are also in widespread use. Compared with portable devices, permanently installed navigation devices in the vehicle provide better positioning quality and thus better route-guidance quality, since additional sensors for distance and direction signals (wheel-speed and yaw sensors) can be evaluated and the antenna can be mounted in a more favorable place for satellite reception. As original equipment, networking with other components is also common, i.e. integration in the operating concept of the vehicle is possible. Voice outputs can be issued via the audio system and muted for telephone calls. The route-guidance information can be displayed in the instrument cluster or in the head-up display and thus in the driver's primary field of vision.

Functions of navigation

Positioning

Satellite positioning system GPS

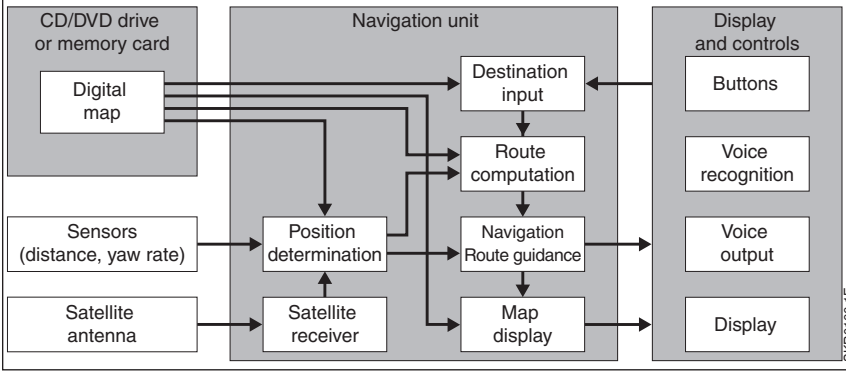
GPS is based on a network of 24 US military satellites which can be used globally for positioning (Figure 2). They circle the Earth in six different orbits at an altitude of approximately 20,000 km in twelve-hourly intervals. They are distributed in such a way that from every point on the Earth always at least four (for the most part up to eight) are visible over the horizon.

The satellites transmit special position, identification and time signals 50 times per second on a carrier frequency of 1.57542 GHz. The satellites each have on board two cesium and two rubidium atomic clocks which vary by less than 20 to 30 ns for the high-precision determination of the transmission time.

Position determination

The signals from the satellites arrive at the vehicle at different times on account of the different propagation times. The position of the receiver is calculated using the trilateration process. When the signals from at least three satellites arrive, the navigation device can calculate its own geographical position in two dimensions (terrestrial longitude and latitude). There is precisely one point which satis-

Figure 1: Navigation system



fies the distance conditions (signal propagation times). When the signals from at least four satellites arrive, this makes possible a three-dimensional position calculation (with altitude). Figure 3 shows this process in simplified form in only two dimensions.

Accuracy
The achievable accuracy is dependent on the position of the satellites that can be received relative to the vehicle. The greater the solid angle of the satellites to the vehicle, the better the possibility of position determination. The achievable accuracy is in the plane approximately 3 to 5 m, and for the determination of height approximately 10 to 20 m.

In deep urban canyons, satellite signals can only be received if the satellites are arranged to a large extent in a line, i.e. in the direction of the street/road. However, the solid angle formed by the satellites is then very small and position determination inaccurate.

Errors in position determination may arise as a result of the reflection of the satellite signals, on metal-coated building fronts, for example.

Travel-direction determination
The direction of travel can be quickly sensed from the differences in the receive frequency of the satellites which are brought about by the Doppler effect. When a car drives towards a satellite, the GPS receiver of the navigation devices sees a higher frequency than the transmit frequency. It receives a lower frequency from satellites behind. This effect is sufficiently great from a driving speed of approximately 30 km/h to determine the direction.

Dead reckoning
Dead reckoning ensures position determination even if no GPS signals can be received, e.g. in tunnels. It adds cyclically recorded distance elements vectorially by magnitude and direction. The speedometer signal transmitted via the CAN bus is used to measure the distance. Changes in direction are recorded by a yaw sensor. In this way, the direction of travel is determined starting out from an absolute direction which was calculated with the GPS signals received last via the Doppler effect.

Map matching
The procedure known as map matching continuously compares the located position with the route on the digital map. In this way, the exact vehicle position can also be shown on the map, even if the located position is inaccurate (caused for example by the lack of a GPS signal and errors in dead reckoning). Route-guidance driving recommendations can thus be output at the best possible location. In addition, sensor errors and cumulative dead-reckoning errors can be compensated.

Figure 2: GPS satellite positioning system
1 to 24 satellites for determining the position of a vehicle.

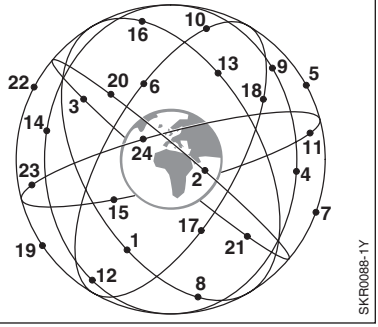
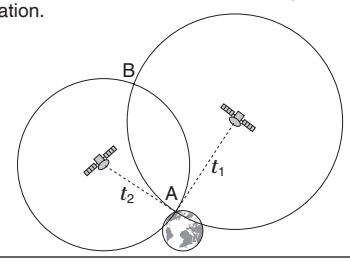


Figure 3: Position determination with GPS
(Simplified two-dimensional view). Where the positions of the satellites are known, the possible reception locations for the measured signal propagation times t_1 and t_2 are on two circles around the satellites. The point of intersection A on the Earth is the sought-after location.



Destination entry
The destination is entered, for example, via the navigation-device panel with buttons, directly via the touchscreen, or by voice input. The user enters all the necessary details from a menu or in response to prompting by acoustically output instructions.

The digital map contains directories so that a destination can be entered as an address. Lists of all known place names are required for this purpose. In turn, all locations are allocated lists containing the names of the stored street names. To pinpoint a destination even further, the user can then also select road/street intersections, or the number of a building.

The destination can be quickly entered via the destination memory, with which already entered destinations can be called up again (e.g. the last destinations used or the destinations saved as favorites).

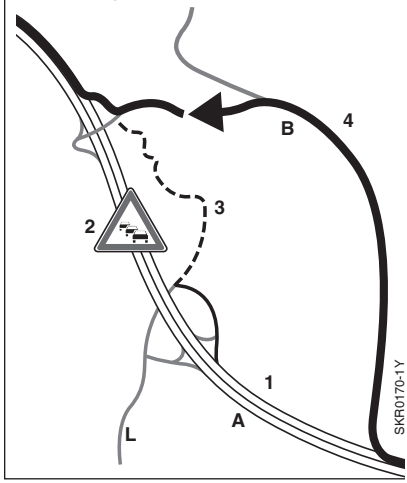
For airports, train stations, filling stations, multi-storey parking garages, etc. there are thematic directories in which these destinations (POI, Points of Interest) are listed. These directories make it possible, for example, to locate a nearby filling station or multi-storey parking garage (Figure 4).

In many navigation systems, destinations can also be marked in the map display (e.g. via the touchscreen).

Figure 4: Example of a map display with POI



Figure 5: Dynamic route guidance
1 Original main route, 2 Congestion,
3 Alternative route estimated by driver,
4 Best alternative route calculated by dynamic route guidance.
A Interstate/expressway,
B National highway,
L Secondary road.



Route computation
Standard computation
Starting from the current location, the navigation device calculates a route leading to the entered destination. The way in which the route is calculated can be adapted to suit the driver's preferences. He or she can stipulate different options such as, for example, optimizing the route
– based on driving time,
– based on an economic mean of driving time and driving distance,
– based on an estimated minimum fuel consumption,
– avoiding interstates/expressways, ferry crossings, or toll roads.

Route recommendations are expected within a few seconds of the destination being entered. A more critical scenario is recomputing the route if the driver departs from the recommended route. The new route recommendations must be supplied before the driver reaches the next intersection or junction.

Dynamic routing
Many radio stations transmit traffic messages not only as spoken text, but also in encoded form. The ALERT-C standard for

the Traffic Message Channel (TMC) is provided for this purpose. TMC content is transmitted via the Radio Data System (RDS) of FM radio.

The encoded messages contain, among other things, the location of a holdup, its extent, the actual length, and the reason for it. A navigation system can receive such an encoded message and determine whether a holdup is situated on a planned route. If this is the case, the route is recalculated, where the held-up section is evaluated with a longer driving time. A new route bypassing the holdup may be obtained (Figure 5). Further route recommendations will follow in accordance with the new route.

The TMC codes required for this are limited to interstates/expressways and major highways.

Route guidance

Route recommendations

Route guidance is performed by comparing the present vehicle position with the computed route. Route recommendations are mainly reproduced acoustically (voice output). This allows the driver to follow the directions without being distracted from the task of driving. Graphics, preferably positioned in the primary field of vision (e.g. in the instrument cluster), support the driver in understanding the directions. These graphics range from simple arrow symbols (Figure 6) through to the display of a map excerpt optimally adapted in size (Figure 4).

Figure 6: Pictogram as example of route guidance



Digital map

Map display

Depending on the system, the map is represented on color displays using scales as of approx. 1:2,000 in 2D, perspective 2D (pseudo 3D), or real 3D. This is helpful for obtaining a general overview of the route in the immediate locality or over a wider area. Additional information such as, for example, bodies of water, railroads and forests provide added orientation.

Digitization

High-precision official maps and satellite and aerial photographs provide the basis for digitizing the data. If the copies are insufficient or not up to date, on-site measurements/surveys are carried out. Digitization is performed manually from map material or satellite and aerial images. Names and classification of objects (e.g. streets, bodies of water, borders) are then integrated into the database.

Specially equipped vehicles drive along roads to record additional traffic-relevant attributes (e.g. one-way streets, passage restrictions, overpasses and underpasses, no turning off at intersections), after which the initial-digitization data are checked on site. The results from these recording vehicles are incorporated into the database and are used to produce digital maps.

Data memory

Because their storage capacity is at least seven times larger, DVDs have now superseded CDs as the storage medium for digital maps. On the other hand, new systems are mainly being supplied with hard disks or SD cards, where the latter clearly dominate in portable systems. These writable storage media open up the additional possibility of the navigation systems being able to adapt to their users' preferences.

Night-vision systems

Applications

Darkness leads to two significant restrictions to a driver's sense of sight. Firstly, despite modern lighting technologies such as xenon and LED headlamps, only a very limited illuminated field of vision is available. For example, the range of a typical low-beam headlamp is just 50 to 60 m. The visibility range can be increased to over 150 m by using high-beam headlamps, this can only rarely be used during night driving, however, due to the tendency to cause glare for oncoming traffic.

Secondly, the capability to recognize objects in the dark by their color and contrast – which differ from those during daylight – is often greatly reduced. Pedestrians wearing dark clothing, for example, are frequently difficult to recognize even when they are within the range of the low beam headlamp. Night-vision systems and cutting-edge headlamp technologies help enormously in this respect to increase road safety.

Camera-based night-vision systems are classified into two different types based on their underlying spectral ranges: far- and near-infrared systems.

Far-infrared systems

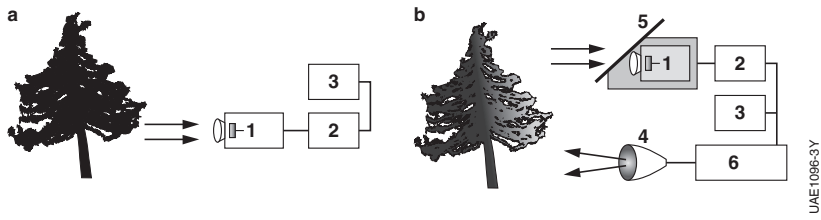
Operating principle

Night-vision systems based on far infrared (FIR) have been used in military applications for years. They use heat radiation in a spectral range between 7 and 12 µm which is relatively far removed from the visible spectral range (hence "far infrared"). They were used for the first time in motor-vehicle applications in 2000 in the USA. They detect with a thermal-imaging camera the heat radiation that is emitted by objects (Figure 1a). These are referred to as "passive" systems since they do not need any additional radiation sources to illuminate objects.

The pyroelectric thermal imaging camera, or microbolometer camera, is only sensitive in the wavelength range of 7 – 12 µm. Because the glass in the windshield is not transparent for these wavelengths and the camera must be located on the outside of the vehicle.

Thermal imaging cameras use germanium lenses, or tex-glass with high amounts germanium and a corresponding transmission band. Currently available cameras have a resolution up to VGA resolution (video graphics array, 640 × 480 pixels). The camera signals are processed in an ECU. The generated video signal is output to a display (in the instrument cluster, in the center

Figure 1: Comparison of far-infrared and near-infrared systems
a) Far-infrared system (7…12 µm),
b) Near-infrared system (780…980 nm).
1 Camera, 2 Electronic control unit (ECU), 3 Display,
4 Infrared headlight, 5 Windshield, 6 Vehicle.



console or in the windshield as a headup display), on which the thermal image can be viewed.

Image display
Hot objects are shown in the image as light contours against dark (cold) surroundings (Figure 2a), where the contrast improves the higher the temperature difference between object and air temperature. However, if a warm object (e.g., a person) is standing in front of a hot tail panel of a tourist bus, he will not be defined by the camera since the sensor is "dazzled".

However, the image display is rather unusual for the viewer since the appearance does not correspond to that of a normal reflected image.

Figure 2: Comparison of the appearance of far-infrared and near-infrared systems
a) Far-infrared system,
b) Near-infrared system.



Near-infrared systems

Operating principle
Near-infrared systems (NIR) are based on infrared radiation between 800 and 1000 nm near the visible spectrum (near infrared). Objects do not emit any radiation in this wavelength range. The area ahead of the vehicle must be illuminated with infrared headlamps (Figure 1b). They are referred to as "active" systems for this reason. An infrared-sensitive camera records the scene, i.e., the infrared radiation reflected by the objects. The image signal is transferred to a control unit that in turn forwards the processed image – as in the far-infrared system – to the display. The driver sees – in both systems – a current image of the road situation (Figure 2b) that plays out like a film.

Functioning principle
Near-infrared night-vision systems exploit the different spectral sensitivities of the human eye and electronic silicon-based images. The spectral sensitivity of the human eye is described by the $V(\lambda)$ curve (Figure 3) and comprises the wavelength range from 380 nm (violet) to 780 nm (red). The maximum sensitivity lies in the range of 550 nm (green). Compared to this, the spectral sensitivity of an imager reaches considerably further into the range of high wavelengths, only ending at around 1,100 nm.

Infrared headlamps
Halogen lamps, like those commonly used for automobile headlamps, have a infrared radiation content. This ranges from the limit of the visible spectrum (380 – 780 nm) up to wavelengths beyond 2,000 nm, with a maximum between 900 nm and 1,000 nm. The upper limit of the useful wavelength when a video camera is used is 1,100 nm, the sensitivity limit of silicon.

Conventional halogen high-beam headlamp models are used as infrared headlamps for this reason. In halogen high-beam headlamp models, visible light is blocked by an additional filter in the light path without significantly reducing the usable infrared part (Figure 3).

The emitted radiation cannot be perceived by the human eye.

The infrared headlamps that are used have a similar range and spatial characteristics to those of conventional high-beam headlamps, which means that the visual range achieved is comparable but without negatively dazzling other road users (Figure 4).

Suitable selection of the filter characteristic curve takes competing requirements into account. The spectral sensitivity of the imager decreases as the wavelength rises. In order to exploit its sensitive range, the filter edge for low wavelengths should lie near to the limit for visible light. On the other hand, in order to prevent non-permitted red impression for front headlamps, the filter edge should be shifted more towards higher wavelengths. Both requirements can be met by carefully selecting the transmission in the visible range, the position of the filter edge, and the edge steepness.

The infrared headlamps are either integrated into the headlamp module or mounted as external modules e.g. in the front apron of the vehicle.

Imagers

The extreme lighting conditions – darkness and very high intensity dynamics, e.g. due to headlamps of oncoming traffic – result in very high requirements for the projection properties of the camera lens, as well as for the imager. The sensitivity to darkness of the imager determines the visual range the system can achieve; the extent of dynamics is deter-

Figure 4: Range of vision of night-vision system

1 Diagram (display),
2 Low beam,
3 Infrared high beam,
4 Range of vision of video camera.

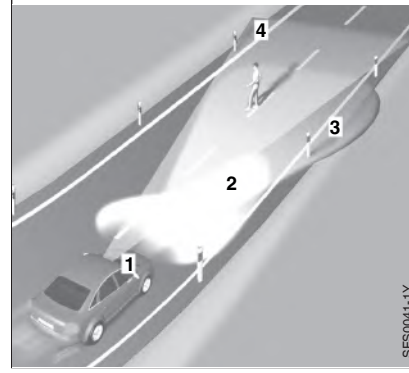
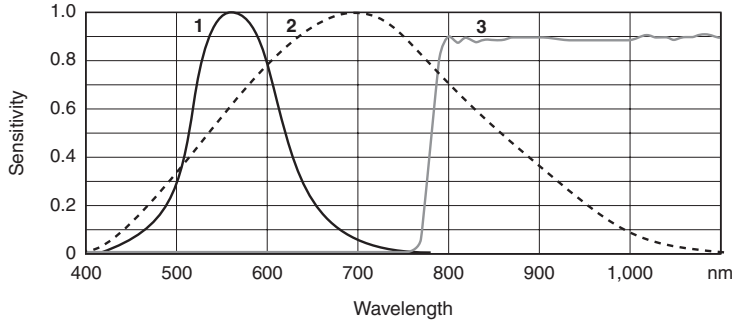


Figure 3: Spectral sensitivity

1 Spectral sensitivity curve $V(\lambda)$ of the human eye (curve for photopic vision),
2 Spectral sensitivity of video camera,
3 Spectral transmission of filter.



mined to a great extent by the resistance to dazzling.

The CCD chip (Charge Coupled Device) features very high sensitivity to darkness, but as a rule it does not achieve the required intensity dynamics without additional measures. In addition, it tends in the case of large brightness differences in the image to override individual, bright image areas. The CMOS chip (Complementary Metal Oxide Semiconductor) has lower sensitivity to darkness, but doe achieve an intensity dynamic of over 100 dB. There needs to be compensation for every pixel of the fixed pattern noise of this technology. Adjustable, nonlinear characteristic curves mean that these imagers can be very easily regulated and adapted to changing lighting conditions.

Lighting up the road scene with radiation in the wavelength range between 700 nm and 1,000 nm (infrared) provides the imager of the camera with a useful signal that depends on the illumination strength and spectral reflective capabilities of the illuminated scene.

Activation strategy

One major aspect of the system is the automatic activation of the infrared headlamps. They emit light of high intensity that the human cannot perceive or can only perceive to a minor degree. To avoid damage to the eyes caused by unperceived exposure at short distances (e.g., mechanics working in the service workshop), the infrared headlamps are only activated automatically when a certain driving speed is reached with the low-beam headlamps switched on. If the speed falls below the threshold, it is automatically deactivated. This excludes occurrence of the critical combinations of long exposure and spatial proximity to the emission source. In addition, the infrared headlamps only light up if the adjacent low beam is also switched on. Through this, the blinking reflex is triggered when the headlamp is looked at, as a result of which the exposure of the eye is greatly reduced.

1st generation night-vision systems

The first system was introduced in 2000 in the American market in Cadillac cars and showed the image of an FIR camera in a head-up display (HUD). This was closely followed by NIR systems in the Japanese market, which also used a head-up display. Due to the high risk of distracting the driver caused by moving images in the primary field of view, these systems were withdrawn from the market after a few years. They could not make the breakthrough in the European market either.

In Japan these first NIR-based systems were initially introduced in 2003 in a simple version in terms of image display and with conventional CCD camera technology. In 2005 an NIR night-vision system with significantly enhanced performance was introduced in the Mercedes S-Class (Night Vision).

In German cars the processed image from the camera (in both far- and near-infrared systems) was displayed in the first systems on a graphics-capable monitor in the instrument cluster (Mercedes) or in the center-console area (BMW). When positioning the monitor in the vehicle, it is important to ensure that the monitor is as close as possible to the windshield and not too far removed from the driver's normal line of sight so that the driver does not have to divert his/her attention from the road and traffic for too long. Ergonomic tests featured the best results in relationship to read-off time and distraction in an arrangement with the monitor in the instrument cluster.

2nd generation night-vision systems

New image-processing procedures enable pedestrians to be reliably classified by reference to their typical outlines (head and shoulder area). This makes it possible not only to warn the driver acoustically but also still to highlight the detected pedestrian on the monitor in order to improve the driver's perception of a dangerous condition. This is particularly important in far-infrared systems, because the image on account of its strange appearance is more difficult to interpret than a near-infrared image.

This feature is important insofar as the majority of fatal accidents involving pedestrians happens at night. In this way, such systems can help to significantly reduce nighttime accidents involving pedestrians. Systems with pedestrian recognition are now fitted as standard in both technologies (near- and far-infrared systems) in German vehicles.

The further development of head-up display technologies offers the additional display of a warning when pedestrians are detected in the primary field of view. Displaying a warning symbol on the windshield represents and effective display format with minimal distractions.

An alternative for simpler systems is a cost-effective optical warning on the base of the windshield that displays the direction of the detected obstacle. These display formats would be a cost-effective alternative for lower classes of vehicles.

3rd generation night-vision systems

New headlamp technologies in the meantime permit the omission of the image display, and as a result the driver is not distracted by looking at the display. A pedestrian detected by the image-processing system can be lit or “flashed” with the aid of a small additional headlamp. This automatically directs the driver's attention towards the pedestrian, who now knows they have been detected.

New pixel-light concepts with numerous high-power LEDs whose light is diffracted by micro-mirrors make it possible to create light cones of almost any shape and facilitate continuous driving on high beam. If a pedestrian or an oncoming vehicle is classified by the imaging system, the light/dark boundary is automatically lowered to such an extent that dazzling no longer occurs.

Parking and maneuvering systems

Application

On virtually all motor vehicles, the bodies have been designed and developed in such a way as to achieve the lowest possible drag-coefficient values in order to reduce fuel consumption. Generally speaking, this trend has resulted in a gentle wedge shape which greatly restricts the driver's rear view when maneuvering. Obstacles are difficult to recognize. This also applies to obstacles beside the vehicle and objects in the driver's blind spot.

Ultrasonic parking aids were initially developed in the past to improve the overview around the vehicle. Thanks to new algorithms and the inclusion of video technology, these systems have undergone further development into maneuvering aids through to (semi-)automated parking systems.

A distinction is made between passive and active systems based on the system configuration. Passive systems warn or inform the driver about hazardous situations, while active systems intervene on a task- and situation-specific basis and steer the vehicle, such as into a parking space.

Ultrasonic parking aid

Parking aids with ultrasonic sensors support parking maneuvers. They monitor an area of approx. 20 cm to 250 cm behind and, if necessary, in front of the vehicle. Obstacles are detected and their distance to the vehicle is indicated by visual or acoustic signals.

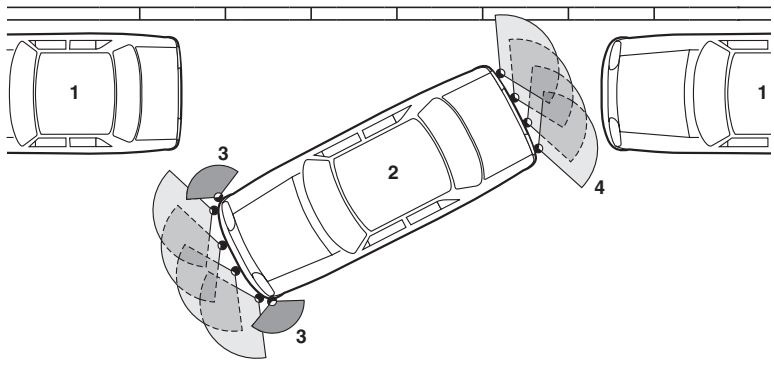
Numerous vehicle manufacturers offer parking aids as optional equipment; in many premium vehicles they are now fitted as standard equipment. Systems which can also be adapted to older vehicles are offered for retrofitting.

System

The parking aid essentially consists of ultrasonic sensors (see Ultrasonic sensor technology), the electronic control unit (ECU), and warning elements. The protection area is determined by the range and number of sensors and by their emission characteristic.

As a rule, vehicles with purely rear protection have four ultrasonic sensors in the rear bumper; some large vehicles, e.g. off-road vehicles (SUVs), also use six sensors. Front-end protection is provided by four

Figure 1: Scanning range of the parking system with rear- and front-end protection
1 Parked vehicle, 2 Parking vehicle, 3 Ultrasonic sensor at the front end,
4 Ultrasonic sensors at the rear end.

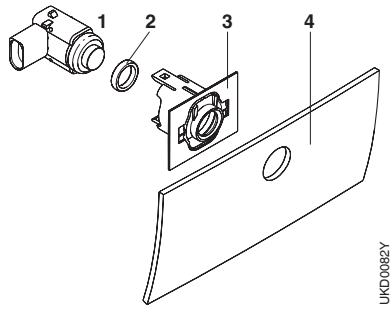


to six additional ultrasonic sensors in the front bumper (Figure 1). For vehicle integration, the installation angle and spacing of the sensors are determined specific to the vehicle. This data is taken into account in the control unit's calculation algorithms. Specifically adapted mounting brackets secure the sensors in their respective positions in the bumper (Figure 2).

The system is activated automatically when reverse gear is engaged. In the case of systems with front-end protection, the system is activated automatically when the vehicle speed drops below a threshold of approximately 15 km/h (manufacturer-specific). During operation, the self-test function ensures continuous monitoring of all system components.

Distance measurement
The ultrasonic sensors emit ultrasonic pulses and receive the sound reflected by an obstacle. The piezoceramic element generates an analog electrical signal, which is evaluated in the sensor. The distance between the vehicle and the detected obstacles is derived from the propagation time between the ultrasonic pulse being emitted and the echo pulses being reflected back from obstacles (see Ultrasonic sensor technology).

Figure 2: Principle of mounting an ultrasonic sensor in the bumper
1 Sensor,
2 Decoupling ring,
3 Installation housing,
4 Bumper.



Electronic control unit (ECU)
The ECU contains a voltage stabilizer for the sensors, a microcontroller, and all the interface circuits needed to adapt the different input and output signals. The software assumes the following functions:
– activating the sensors and receiving the echo,
– evaluating the propagation time and calculating the distance to the obstacle,
– activating the warning elements,
– evaluating the input signals from the vehicle (e.g. signal for engaged reverse gear),
– monitoring the system components including fault storage,
– providing the diagnosis function.

Warning elements
The warning elements display the distance from an obstacle. Their design is specific to the vehicle, and they usually provide for a combination of acoustic signal and optical display. Primarily LEDs, but also LCDs are currently used for optical displays.

In the case of vehicles with a display monitor in the center-console area or in the instrument cluster, the vehicle can be shown together with the obstacles detected by the system, e.g. from a bird's eye view. This generally improves the driver's view of the situation.

Ultrasonic parking assistant

The parking assistant is based on the ultrasonic parking aid and is designed along evolutionary lines in stages. Each stage is a self-contained function.

All the ultrasonic systems shown in the following are already in series production.

Parallel parking

Information-providing parking assistant

After the system is activated by the driver, an ultrasonic sensor mounted on each side of the vehicle measures the length and depth of the parking gap as the vehicle drives past the gap (Figure 3). The length is obtained by evaluating the ESC speed signal (ESC, Electronic Stability Control, driving-dynamics control). The parking assistant then issues the driver with a signal as to whether the parking gap is long enough, tight or too short. If there is an obstacle in the parking space, this is rejected. This requires the above-mentioned sensors with a range of up to approximately 4.5 to 5.5 m.

Once the parking space has been measured, the determined geometry of the environment can be used to ascertain the optimal trajectory for the parking process. The parking assistant requires the signals from the steering-angle sensor (from the driving-dynamics control, ESC) for this purpose. During the parking process, the system can now give the driver a recommendation regarding how he or she should optimally turn the steering wheel in order to park in the space as smoothly as possible. During the parking operation, the trajectory is continuously recalculated and shown in a display. Figure 4a shows an example of a driving recommendation from the system.

Steering parking assistant

The next evolutionary stage is a system with automatic steering activation. A prerequisite is that the vehicle must have electrically activated power steering.

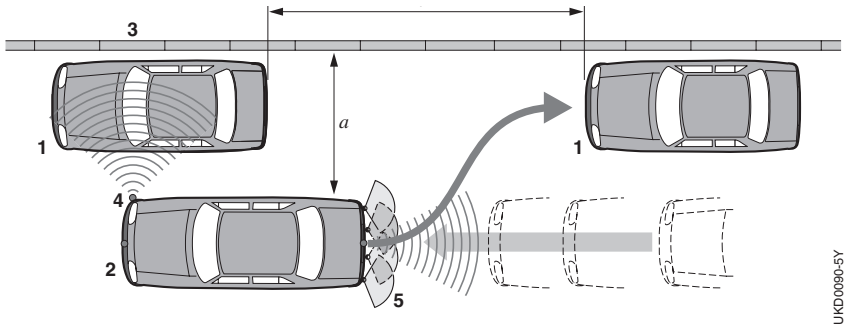
As before, the driver has to inform the system of his or her desire to park by pressing a button. After the system measures the parking space, the driver is notified whether the length of the parking space is sufficient. This can be accomplished for example by a corresponding graphics display in the instrument cluster (Figure 4b). Now the driver simply needs to put the vehicle in reverse and control the vehicle's linear motion (by braking and accelerating) – the parking assistant takes over the steering. When the parking process is done, the driver is notified accordingly.

One-step and multi-step parking

In the case of both the information-providing and the steering parking assistant, the system decides whether to park in one step or multiple steps, based on the length of the parking space. One-step parking is always selected if the parking

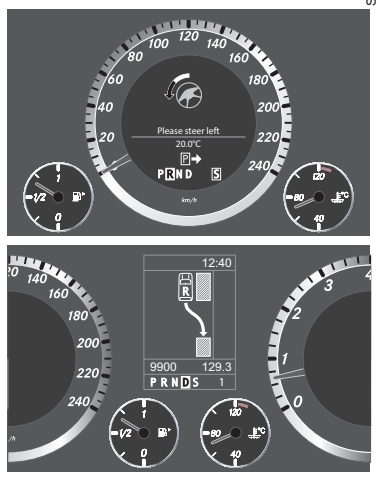
Figure 3: Parking assistant

1 Parked vehicle, 2 Parking vehicle, 3 Parking space boundary (e.g., curb), 4 Side-detecting ultrasonic sensor, 5 Ultrasonic sensor in the rear area.
a Measured depth of the parking space, *l* Length of the parking space.



space is long enough. If the parking space is short, the parking assistant decides on multi-step parking, based on the length of the space and the vehicle's possible steering radius. Then the system prompts the driver to shift gears accordingly.

Figure 4: Display of the parking assistant of the multifunctional display in an instrument cluster
a) Display of the information-providing parking assistant,
b) Display of the steering parking assistant.



Perpendicular parking

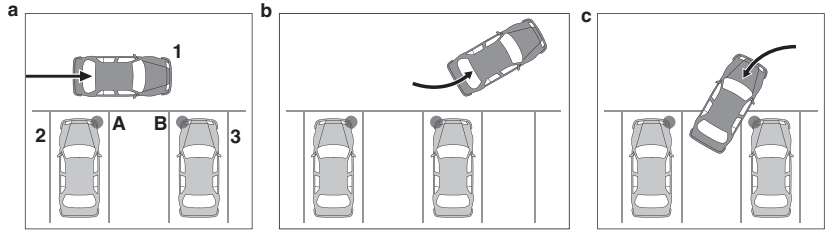
Semi-automated parking is not limited to the parallel parking functions described, but also permits parking in spaces perpendicular to the direction of travel (Cross-Parking Assist, Figure 5).

As with parallel parking, the length of the parking space is measured as you drive by it. To do this, the corner of the vehicle on the left (marked as corner point A in Figure 5) is taken as the initial reference point for the parking maneuver. If there is also a vehicle to the right of the parking space, its corner (corner point B) is also taken into account.

Before beginning the parking process, the vehicle moves forward until it can calculate a parking trajectory. Then, after reverse gear is engaged, the vehicle parks along the trajectory, which is continuously recalculated and corrected if necessary. The signals from the rear ultrasonic sensors are continuously factored into the parking maneuver. Due to the limited detection range of the ultrasonic sensors, the driver is responsible for making sure the entire length of the parking space is free of obstacles.

As with the previously described parking assistants for parallel parking, for this function there are also information-providing and – if the vehicle is equipped with electric power steering – steering system versions.

Figure 5: Perpendicular parking
a) Detection corner point A,
b) Preparation for parking,
c) Parking in the parking space.
1 Parking vehicle, 2 Left parked vehicle, 3 Right parked vehicle.
A Right corner point of the left parked car,
B Left corner point of the right parked car.



Assistant for leaving a parking space
In principle, the parking aid can also be used for leaving parallel parking spaces. To do so, the system first moves the vehicle into an appropriate starting position, so that the maneuver can be done in one step. The driver is responsible for paying attention to the traffic and accelerating or braking at the right moment. Simply touching the steering wheel turns off the function.

Maneuvering assistant
Traces of paint on parking-garage pillars and walls testify to frequent, sometimes costly collisions with vehicles. The system for warning about objects next to the vehicle helps prevent that.

Ultrasonic sensors mounted on the front of the vehicle at the sides detect permanent obstacles in the vehicle surroundings at a slow driving speed (up to 30 km/h) (Figure 6). The maneuvering assistant saves their positions, so that these are known even after the obstacles are no longer in the scanning range of the sensors. The side distance to the saved obstacles is continuously calculated using the steering angle sensor signal. If a risk of collision is detected, the driver receives an audible warning (side distance warning). A bird's-eye view of the situation can also be shown on the central display. This system was introduced to serial production in 2013.

Video systems

Rearview camera
Rearview cameras were already first introduced some time ago in Japan, but did not become prevalent in Europe due to the heavily distorted picture of the super-wide-angle lens. The breakthrough in Europe didn't come until there were systems to correct the image distortion.

Today's systems have a wide-angle camera installed in the rear end of the vehicle, usually in the recess of the handle to the trunk. The camera image appears in the center console display when driving in reverse. Newer systems with additional lines to help estimate the distance and display the predicted lane make it easier for the driver to maneuver (Figure 7).

The camera is intended only to provide information and does not provide the driver with any warnings.

Front-end camera
Visibility all around the front of the vehicle can also be provided by a video camera with a 180° wide-angle lens. It also provides a view of traffic from the sides, for example, when pulling out of a driveway (Figure 8). Since the driver is seated farther back in the vehicle, the view from there is not as good as from the camera installed in the front of the vehicle.

Figure 6: Maneuvering assistant
1 Vehicle driving into a parking garage,
2 Parked vehicles,
3 Obstacle.

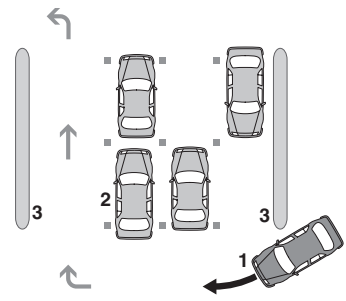


Figure 7: Driver assistance with rearview camera

Video display with additional information displayed. The distance information is calculated from the optical parameters of the single video camera.
1 Trajectory without turning the steering wheel,
2 Trajectory calculated from the current steering angle.

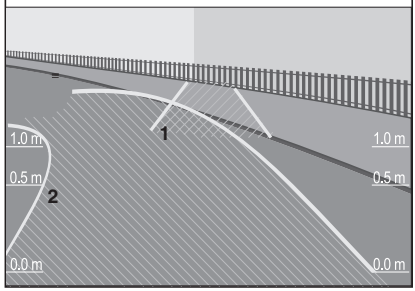
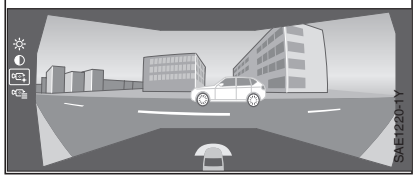


Figure 8: Driver assistance with front camera

View of the main road when driving out of a private driveway.



360° all-around visibility

The whole area around the vehicle can be detected using four super-wide-angle cameras installed around the vehicle (top-view system, Figure 9). The fields of view of the individual cameras overlap in the shaded areas. The four individual images are converted into one overall image using the stitching method. Projecting the four images into an imaginary bowl structure makes it possible to provide the driver with a bird's-eye view and a view from different directions of his or her own vehicle in the display.

Figure 10: Video image with distance information of the ultrasonic sensor

1 Obstacle (e.g., wall),
2 Distance assistance lines on the 50 cm grid, calculated from measured values of the ultrasonic sensors,
3 Trajectory at maximum steering angle.

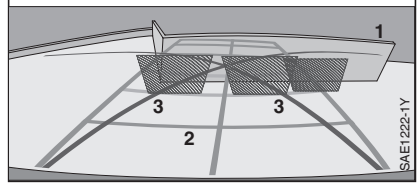
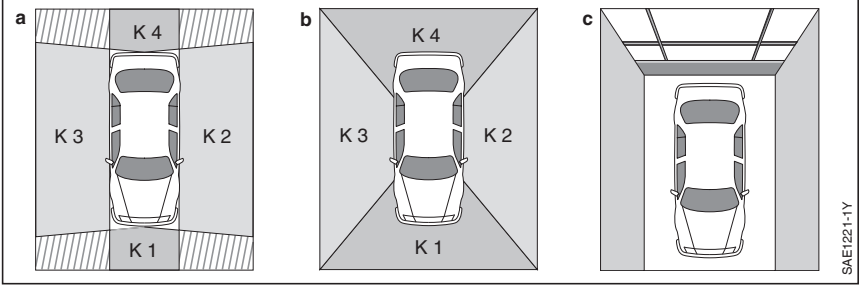


Figure 9: 360° all-around visibility with four cameras

a) Detection range of the cameras (shaded areas indicate where they overlap),
b) Strictly defined image borders for the top view image,
c) 360° all-around visibility from a bird's-eye view (parking garage example).
K1 to K4 Cameras on the front, rear and sides of the vehicle.



Fusion of video and ultrasound technology
Further improvements can be achieved through data fusion of the ultrasound system's signals with the video camera's signals. For example, in addition to the lane information, distance information from the ultrasound system can be displayed in the image from the mono rearview camera as well, which cannot be used for measuring distance (Figure 10). By doing so, the driver receives even more meaningful information about the area by the rear of the vehicle.

More recent development steps
Further functions and options have been created through the further development of semi-automated parking by using intelligent image-processing algorithms (see Computer vision). The combination of intelligent camera systems with ultrasonic sensors form the basis for the automated parking and maneuvering assistance systems outlined in the following.

Parking maneuver assistant (park maneuver control)
Further development of the steering parking assistant has the vehicle now controlling linear motion as well. To initiate the parking maneuver, the driver needs only to put the gear in reverse; the system handles the rest. This system was launched into series production in 2014.

Remote-controlled parking assistant (remote park assist)
The driver drives past a parking space and receives a signal from the parking assistant is the space is sufficiently long. The driver gets out of the car and taps the "Parking" function on their smartphone. As long as the driver leaves their finger on the smartphone screen the car drives along the previously calculated trajectory into the parking space. If during the parking maneuver a critical situation arises (e.g., pedestrian crosses the trajectory), the driver simply has to take their finger off the screen and the car will stop immediately.

Home-zone park assist
In parking garages or parking lots with reserved parking spaces in the home environment it will in future be possible to teach the car a parking maneuver from a defined parked position to the driver's own parking space by driving the route and maneuvering the car into the parking space. The driver can then leave the car in the parked position and start the car by actuating a remote control as described above. The car then automatically drives the taught course to the parking space and parks up there. If the driver wants to use their car again, they can fetch it again from the garage with the same remote control. The car will then stop in the original parked position.

Highly automated, remote-controlled parking (remote park pilot)
This function is used above all for maneuvering into and out of tight garages or parking spaces. Because it is in principle possible to park a car without the need for a driver, the user can start the parking operation in front of the garage with a remote control, an intelligent remote key, or their smartphone by touching an operating element. The car then drives under automated guidance into the garage. The driver does not need to be present for this. This function calls for very robust environment recognition and high safety requirements.

This system was launched into series production in 2016.

Autonomous parking (automated valet parking)
A car that parks autonomously was presented to the public for the first time in 2007 as part of the American "Urban Challenge", a competition for automated driving in an urban environment. Here the car was parked at the entrance to a parking lot. The car found a free parking space itself and parked in it.

With this new function too the car is parked in a predefined drop-off area in front of a parking garage or a basement garage. For the subsequent parking maneuver the driver gets out of the car and selects the “Find parking space” function on their smartphone. The the car drives under fully automated guidance into the parking garage, finds a parking space itself, and parks in it. The same operating procedure can be used to fetch the car from the garage. For driving and maneuvering the car uses its own on-board sensor technology (ultrasonic and video-based systems) and signals from an infrastructure of the parking garage (see also Automated valet parking).

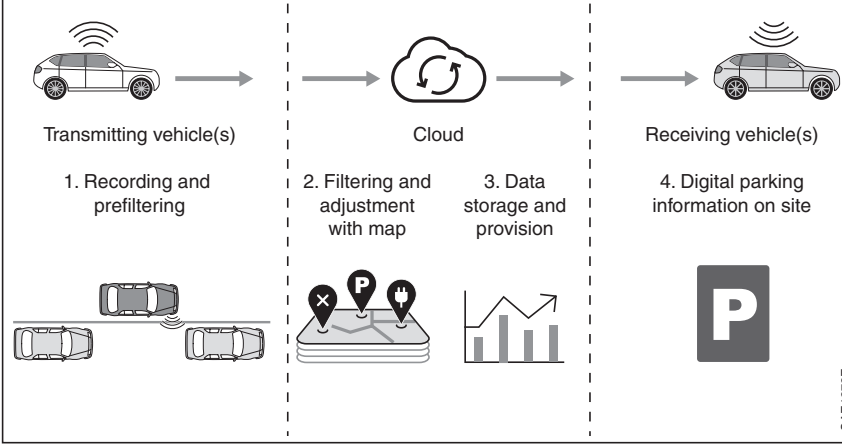
Such systems can have a significant influence on the infrastructure of future parking garages: The vehicles can be parked much more tightly since large spaces do not have to be kept for getting in and out. The storey height can be reduced. Long-stay parked vehicles can be “blocked in”.

Networking with the “Cloud”, Connected Parking

Looking for a parking space involves driving unnecessarily long distances. The resulting fuel consumption and exhaust emissions add up to a huge environmental impact.

Remedial action can in future be taken by a system which actively supports the driver in looking for a parking space. This system is based on the interaction between vehicles and infrastructure (V2I, Vehicle to Infrastructure Communication, Figure 11). Vehicles which are situated in an area with parking spaces (e.g., in an urban area) detect the geographical location of available parking spaces and their length with their on-board sensor technology and transmit this information to a “Cloud”. In the Cloud the data are filtered and the parking position is inserted into a map. All the relevant information is stored and made available to possible users. As shown in Figure 11 by way of example, the driver receives information on occupied and vacant parking spaces and parking paces with a charging station. Vehicles looking for a parking space can download the information from their surroundings from the Cloud and receive route-direction information to a neighboring vacant parking space or to a parking space as close as possible to the input destination in the navigation system.

Figure 11: Parking-space search with Vehicle to Infrastructure Communication



Adaptive Cruise Control

Function

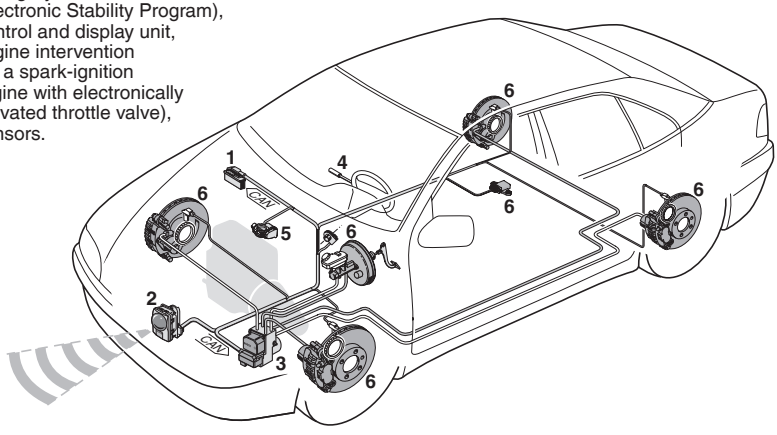
Like the basic cruise-control system that has been available as a standard feature for quite some time, Adaptive Cruise Control (ACC, also known as Tempomat) can be categorized as a driver-assistance system. The vehicle-speed controller regulates the driving speed to the desired speed set by the driver using the operating unit of the cruise control. In addition to this function, adaptive cruise control senses the distance and the relative speed for the vehicle in front and uses this information with additional data from the driver's own vehicle (e.g. steering angle, yaw rate) to regulate the time gap between the vehicles. It thus adapts the speed to the vehicle ahead and maintains a safe distance. The adaptive cruise control is equipped with a long-range radar to detect vehicles that are traveling in front in the same lane or also obstacles that are moving in the sensor's sensing range, and necessitate that the vehicle be braked (Figure 1).

Design and function

Adaptive cruise control is a convenience system that relieves the driver of routine tasks, but not of his/her responsibility to maintain control of the vehicle. This is why the driver can override this function at any time through intervention or deactivation (e.g. by operating the accelerator or brake pedal).

Ranging sensor
Current ACC systems for the most part have a radar sensor which operates in a frequency range of 76 to 77 GHz (see Radar sensors). The radar beams emitted by the radar sensor are reflected by vehicles ahead and analyzed with regard to propagation time, Doppler shift, and amplitude ratio. These values are used to calculate the distance, speed, and angle position relative to the vehicles ahead. The evaluation and control electronics (radar-sensor check unit) are integrated in the sensor housing. They receive and transmit data via a CAN data bus from

Figure 1: Adaptive cruise control for a passenger car
1 Engine control unit,
2 Radar-sensor control unit,
3 Active brake intervention via driving-dynamics control (Electronic Stability Program),
4 Control and display unit,
5 Engine intervention (on a spark-ignition engine with electronically activated throttle valve),
6 Sensors.



or to other electronic control units that influence the engine torque and brakes (Figure 2).

There are also ACC systems that work with laser beams in the infrared range (see Lidar, Light Detection and Ranging). The functioning principle is similar, whereby the use of an optical beam means that restrictions compared to the radar systems must be accepted in poor weather conditions (fog, rain, snow) (see e.g. [5]).

Course setting

To ensure reliable ACC operation no matter what the situation – e.g. also on curves/bends – it is essential that the vehicles traveling ahead can be allocated to the correct lane(s). For this purpose, the information from the driving-dynamics control (Electronic Stability Program) sensor system is evaluated with regard to the ACC-equipped vehicle's own cornering status (yaw rate, steering angle, wheel speeds, and lateral acceleration).

Setting options

The driver inputs the required speed and the required time gap; the time gap available to the driver ranges from 1 to 2 s. The time gap to the vehicle in front is calculated from the radar signals and compared with the required time gap specified by the driver. If the time gap is shorter than the required gap, the ACC system responds in a manner appropriate to the traffic situation by initially reducing engine torque, and only if necessary by automatically braking the vehicle. If the required time gap is exceeded, the vehicle accelerates until either the speed of the vehicle ahead or the desired speed set by the driver is reached.

Engine intervention

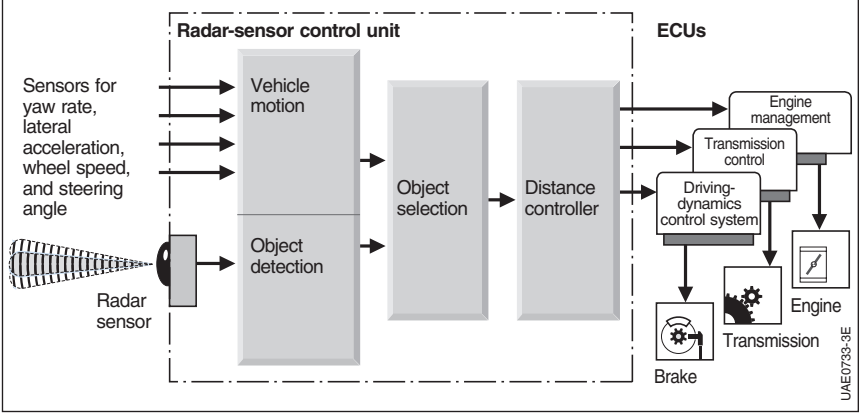
Cruise control is run via the electronic engine-performance control system (e.g. Motronic, electronic diesel engine control). This system allows the vehicle to be accelerated to the required speed or, if an obstacle appears, to be decelerated by reducing the drive torque.

Brake intervention

If the deceleration caused by engine control intervention is insufficient, the vehicle must be braked. For passenger cars the driving-dynamics control (Electronic Stability Program) is required for this purpose; this can perform a brake intervention. For commercial vehicles, an electronic braking system (EBS) is sufficient. Usually, this also involves the available retarder or engine brake for wear-free deceleration.

Due to the configuration of ACC as a comfort system, the deceleration calcu-

Figure 2: Basic structure of ACC system



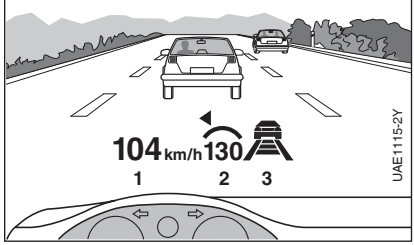
lated by the controller is limited nowadays to approx. 2 to 3 m/s^2. Should this be insufficient due to the current traffic situation (e.g. vehicles in front braking sharply), a visual and audible signal is issued to the driver indicating that he or she must assume control. The driver must then initiate the corresponding braking deceleration via the standard brake. The ACC does not include safety functions such as emergency braking.

If necessary, with ACC activated, the stabilizing systems Antilock Brake System or driving-dynamics control are activated in the usual manner. Depending on the parameterization of ACC, stabilizing interventions from the Antilock Brake System or driving-dynamics control cause ACC to shut off.

Display
The driver must be provided with at least the following information:
– Display of the desired speed
– Indication of the switch-on status
– Indication of the required time gap selected by the driver
– Indication of the follow-up mode, which informs the driver as to whether the system is controlling the distance to a detected target object or not.

The information can be displayed, for example, in the instrument cluster or in a head-up display (Figure 3).

Figure 3: Information shown in the head-up display
1 Present speed,
2 Desired speed,
3 ACC status (Adaptive Cruise Control).



Control algorithms

Control modules
As a general principle, the control system for cars as well as for trucks consists of three control modules.

Cruise control
If the radar sensor has not detected any vehicles in front, the system maintains the vehicle's speed at the cruising speed set by the driver.

Tracking control
The radar sensor has detected vehicles in front. Control essentially maintains the time gap to the nearest vehicle at a constant setting.

Cornering control
When negotiating tight bends, the radar sensor can "lose sight" of the vehicle in front because of the limited width of its field of vision. Until the vehicle comes in sight of the radar again, or until the system is switched to normal cruise control, special measures come into effect. Depending on the vehicle manufacturer, the speed, for example, is kept constant, current lateral acceleration is adapted or ACC is switched off.

Object detection and lane allocation
The central task of the radar sensor and its integrated electronics is to detect objects and allocate them either to the same lane as the one on which the truck is traveling, or to a different lane. This lane allocation requires both exact detection of vehicles ahead (high angle resolution and accuracy) and exact knowledge of the vehicle's own movement. The decision as to which of the detected objects is used as the reference for adaptive cruise control is essentially based on a comparison between the positions and motion of the detected objects and the motion of the system's own vehicle. Here, in particular in the case of commercial vehicles, it is not always necessarily the next vehicle ahead that is selected. Under certain circumstance, it makes better sense to use the next vehicle but one, e.g. when a passenger car filters in from

the slip lane and continues rapid acceleration to change quickly to the fast lane.

Electronic structure
Apart from the data transmitted by sensors (Figure 2), ACC requires additional data from the engine, retarder (for commercial vehicles), transmission and driving-dynamics control units, and which are transmitted via the CAN data bus. Conversely, these control units convert the accelerations requested by ACC into drive and braking torques. The coordination of the actuators (e.g., distribution of the required braking torque to the available brakes) can take place both in the driving-dynamics control unit and in a guidance computer or in the ACC control unit itself.

Adjustment
The radar sensor is fitted in the frontal area of the vehicle. Its radar lobe scans are aligned relative to the vehicle longitudinal axis. This is done by means of the adjustment screws in the mounting area of the sensor. If it is moved out of alignment by physical force, e.g. deformation of the mounting due to accident damage or any other effect, realignment must be carried out. Small degrees of misalignment are automatically corrected by the permanently active alignment routines implemented in the software. If manual realignment is required, this is indicated to the driver.

Area of application and functional expansions

Use in passenger cars
The use of Adaptive Cruise Control is bound to the availability of driving-dynamics control. This is a requirement for active brake intervention without action on the part of the driver.

Use in commercial vehicles
Requirements
Some of the general conditions for a commercial vehicle ACC differ significantly to those for a passenger car ACC:
– Regulation of braking and acceleration must take account of the widely variable parameters of payload and engine size.
– Overtaking and merging maneuvers follow a dynamic pattern that is slower on trucks than on cars, and therefore result in different control requirements and settings.
– The control dynamics must be capable of handling a situation where several vehicles equipped with ACC are driving in a convoy (trucks drive in a line at nearly the same speed).
– A degree of simplification is provided by the fact that trucks have a more limited speed range than cars.
– The commercial deployment of trucks means that the focus during configuration is placed more on economy than on sports and comfort. Fuel consumption and wear must be at least as good with ACC as with an average driver.

As a general principle, the same type of ACC system can be used for buses, trucks and tractor units. The requirements with regard to drive and braking systems, manual, semi-automatic or fully automatic transmissions are also very similar in buses and trucks. Only with the configuration of the system are there heightened requirements in buses with regard to comfort.

System design
The system architecture of ACC for commercial vehicles is barely different at all from that of ACC for passenger cars. However, the range of the sensors can be shorter owing to the lower speeds involved. The environmental and service-life requirements imposed on the components in this respect are much higher.

ACC control is configured in commercial vehicles for optimum consumption and wear performance. Thus, for example, the service-brake system is activated only when the available retarder is no longer able to decelerate the vehicle in the manner requested by the ACC controller. Brake wear caused by ACC is therefore for the most part at the same level as that caused by an average driver. The same applies to fuel consumption, in which case a positive effect compared with an average driver can also be discerned thanks to ACC.

As well as these ACC systems, there are for commercial vehicles also predictive ACC systems which, on the basis of route-topology information that is learned or stored in maps, optimize the driving speed in such a way as to achieve the lowest possible fuel consumption. The systems achieve this for example by no longer accelerating ahead of humps in the road or on downhill stretches with a subsequent uphill gradient by better utilizing the kinetic energy through increased speed. The proven savings are, even when compared with good drivers, in the range of 5 to 10 %.

In some countries, e.g., the USA and Canada, there are even ACC systems which on account of the simpler brake architecture influence solely the engine torque or the continuous braking torque. The driver must therefore intervene manually more often.

ACCplus
First generation ACC systems had limitations in their range of functions resulting from the limited functionality of sensors and actuators. The limited object-sensing range and the limited horizontal resolution capacity only permitted operation above 30 km/h. Therefore ACC could not initially be used down to low speeds and at a standstill. The ACCplus system (in series production since 2009) has a further-developed radar sensor with wider angular coverage (±15°) and improved detection properties, and thus permits braking to a standstill and restarting by driver intervention. Pressing the accelerator pedal within a predetermined time limit is sufficient to reactivate the system.

Thanks to its greater reliability in destination selection and its even better ability to detect objects at close range, ACC can now also be used in traffic-jam situations.

ACC with low-speed following
During low-speed following (ACC Stop & GO, ACC LSF, Low Speed Following), the data from the long-range radar sensor is combined with the data from the medium or short-range sensors (short-range radar sensors or ultrasonic sensors) in order to provide an exact measurement of objects in front of the entire vehicle.

It operates in a speed range of 0 to 200 km/h, brakes the vehicle to a standstill, and automatically restarts within a predetermined time limit.

Sensor-data fusion with a video camera
Object measurement and object classification can be performed by means of sensor-data fusion with a video camera. This enables robust control over the vehicle with regard to stationary objects.

By combining with video sensor technology, it will be possible for low-speed following to perform full linear guidance at all speed ranges and also in urban traffic in the future (FSR, full speed range). ACC has also provided the basis for developing assistance systems which automatically intervene in critical driving situations to avoid accidents or to mitigate the consequences of accidents. Further developments are geared towards automatic evasive maneuvering by automatic steering intervention.

Current ongoing developments

Electronic horizon

The quality of the ACC function is being improved based on digital maps. The navigation system determines the position of the vehicle and estimates the expected path that the vehicle will travel (most probable path, MPP). A wide variety of predictive information is provided along this path by accessing digital navigation maps (electronic horizon). Examples of this include the road class, the geometric shape of the road, expressway exits and on ramps, speed limits and the gradient of the road. Manufacturer-specific protocols or the ADASIS standard [4] using CAN bus are used to provide the electronic horizon to the ECU.

The ACC system can access this information to improve its function. For example, the predictive course of the road can be used to more reliably assign oncoming radar specifically to a lane (reducing adjacent lane interference). The acceleration of the vehicle can be prevented when approaching an expressway exit if an oncoming radar object is lost when switching to the deceleration lane. The set speed can be reduced automatically before coming to a very tight curve.

In the future, the increased enhancement of the digital map with data and improvement in data quality will enable innovative assistance functions in combination with the ACC system, including predictive, fully automated specification of the set speed. The driver no longer has to enter the intended speed manually; the ACC system determines it automatically by communicating with the navigation system.

This allows the ACC system to implement predictive driving strategies in combination with the engine control unit, which minimize CO_2 emissions. If the ACC "sees" a speed limit or curve in advance via the electronic horizon, the vehicle can roll or coast to it with optimal fuel consumption. The gradient of the road is also taken into account when initiating the rolling or coasting procedure.

Networking with the cloud

The higher the ACC function's degree of automation, the greater the requirements are on the timeliness, precision and reliability of the data in the electronic horizon. To increase the data quality of the digital map, the vehicle itself will contribute to recording street environment data using its sensors. The function: Information such as traffic signs or the cornering speed are recorded by the vehicle's sensors, supplemented with data from the navigation system (e. g. position) and sent to the server (cloud) via mobile communications. In the central server, the incoming signals from many vehicles are processed, compiled and prepared for retrieval. The server information can be requested by vehicles with navigation systems. This ensures that the most recent information is always available. The incoming data from the server is applied to the local navigation map as new knowledge and provided to the ACC system immediately via the electronic horizon.

References

[1] H. Winner; S. Hakuli; F. Lotz; C. Singer: Handbuch Fahrerassistenzsysteme. 3rd Ed., Verlag Springer Vieweg, 2015.
[2] H. Wallentowitz, K. Reif: Handbuch Kraftfahrzeugelektronik. 2nd Ed., Vieweg+ Teubner Verlag, 2010.
[3] H.-H. Braess, U. Seiffert: Handbuch Kraftfahrzeugtechnik. 7th Ed., Verlag Springer Vieweg, 2013.
[4] C. Ress et al.: ADASIS PROTOCOL FOR ADVANCED IN-VEHICLE APPLICATIONS, ADASIS-Forum (http://durekovic.com/publications/documents/ADASISv2%20ITS%20NY%20Paper%20Final.pdf).
[5] K. Reif: Automobilelektronik. 5th Ed., Verlag Springer Vieweg, 2014.

Information and warning systems

Rearview-camera system

Function

Modern vehicles often offer the driver only a limited view of the vehicle's immediate surroundings. Increasingly smaller side and rear windows, combined with a vehicle shape that is strongly influenced by aerodynamics and pedestrian protection, is making safe and precise maneuvering difficult. The rearview-camera system assists the driver when reversing by displaying a real-time image of the rear driving area in the radio or navigation-system display. The rearview-camera system offers the driver when reversing an unobstructed view of the vehicle's immediate surroundings (Figure 1).

Rearview cameras were already first introduced some time ago in Japan, but did not become prevalent in Europe due to the heavily distorted picture of the super-wide-angle lens. The breakthrough in Europe didn't come until there were systems to correct the image distortion.

Operating principle of the rearview-camera system

The near-range camera in the vehicle's rear automatically becomes active as soon as the reverse gear is engaged. The camera image displays the area behind the vehicle (detection range up to 180°, Figure 2). This enables the driver to determine in real time whether the way is clear. Optionally, the vehicle's path can be shown dynamically on the camera image with the aid of colored lines. These display the vehicle's trajectory according to the current steering-wheel position and indicate when the steering wheel must be turned. This makes accurate maneuvering possible.

In combination with the ultrasonic-based parking aid, for example, distances measured by the ultrasonic sensors are displayed as colored bars in the camera image. This means that all relevant information is available to the driver at a glance.

System benefits for the car driver

- The rearview-camera system allows for safe and comfortable parking and maneuvering,
- it allows for quicker and easier parking maneuvers,
- it is easy to use,
- the display of the surroundings is easy to understand thanks to intelligent image processing (no fish-eye effect),
- it reduces the risk of material damage and the associated costs.

Figure 1: Driver assistance with rearview camera

Video display with additional information displayed. The distance information is calculated from the optical parameters of the single video camera.
1 Trajectory without turning the steering wheel,
2 Trajectory calculated from the current steering angle.

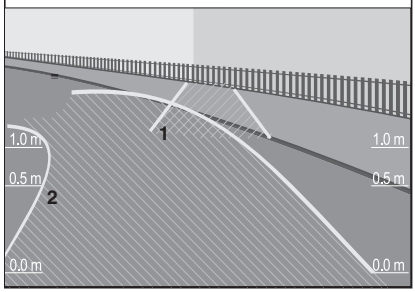
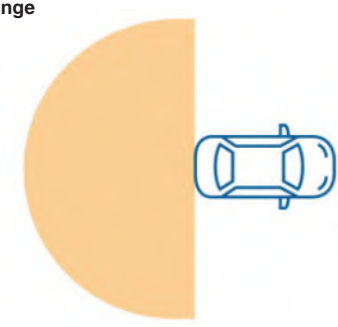


Figure 2: Reversing-camera detection range



Multi-camera system

Features
Parking and maneuvering are among the activities most likely to be associated with accidents, so every centimeter counts. The multi-camera system offers the driver a 360° surround view to ensure easy, safe, and comfortable parking and maneuvering. The option to switch between perspectives makes it possible to line up precisely with lines, curbs, and walls. Safe maneuvering is possible through better orientation and visibility of the vehicle's complete surroundings, including obstacles and curbs.

Properties of the multi-camera system
The 3D surround view of the multi-camera system offers various advantages thanks to its highly complex video-based object detection. The 360° view (Figure 3) and the automatic, real-time tracking shot always provide the driver seamlessly with the best perspective. In addition, the multi-camera system generates a realistic, animated 3D vehicle model that provides the driver with an impressive, detailed image of the vehicle's surroundings. Moreover, the 3D visualization automatically adapts while maneuvering to an environment model based on sensor-data fusion to provide the best view of critical situations. This makes automated parking and parking-assist functions easier and safer.

The modular approach of this multi-camera system covers the various performance requirements with only one ECU, near-range cameras with a resolution of one to two megapixels, and optional ultrasonic sensors. Light system variants integrate two to four near-range cameras, depending on the resolution.

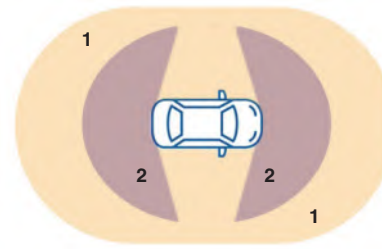
Automated and assisted parking functions can be robustly realized with the aid of the multi-camera system since sensor-data fusion increases the understanding of the vehicle's surroundings and thus improves parking functions like the (remote-controlled) parking assistant or the home-zone parking assistant. The multi-camera system also supports the driver when parking and maneuvering with a trailer. It simplifies coupling the trailer by providing a specific view of the trailer tow hitch and supports the driver when reversing with the trailer by taking over lateral guidance of the vehicle.

Operating principle of the multi-camera system
The multi-camera system consists of four near-range cameras. One faces forward, another rearward, while the two side cameras are integrated in the exterior mirrors. Each with an aperture angle of 190°, they capture the vehicle's complete surroundings. The system also includes an ECU, which merges the camera images to form a 360° view and – if desired – combines them with distance information from ultrasonic sensors.

Using these data, the system generates a dynamic 3D wireframe around the virtual vehicle model; the shape of that frame is adjusted to match to match the changing surroundings. The real-time images from the near-range cameras are projected onto the wireframe. In contrast to 2D systems, this method allows objects that rise up to be represented more realistically.

A new technical feature of the multi-camera system is the seamless, distortion-free merging of the camera images in real time: Depending on the perspective, the system dynamically adjusts the borders of the images. The result is a homogeneous image with no distracting frames

Figure 3: 360° surroundings detection with the multi-camera system with four near-range cameras
1 Detection range of the front camera, rearview camera, and cameras in the exterior mirrors,
2 Detection range of the optionally used ultrasonic sensors.



or overlaps. Objects remain optimally visible at all times.

The multi-camera system offers different modes and perspectives. It supports the driver by affording them an exact view of the vehicle's surroundings which are relevant at that very moment to enable the drive to maneuver and park safely and comfortably. The system uses 3D visualization techniques which are also used in cinema films for a highly realistic image.

System benefits for the car driver

– The multi-camera system facilitates more precise parking and maneuvering the visible of the vehicle's complete surroundings including lines and obstacles,
– different perspectives make it possible to line up precisely with lines, curbs, and walls,
– it helps to avoid collisions and reduce parking stress.

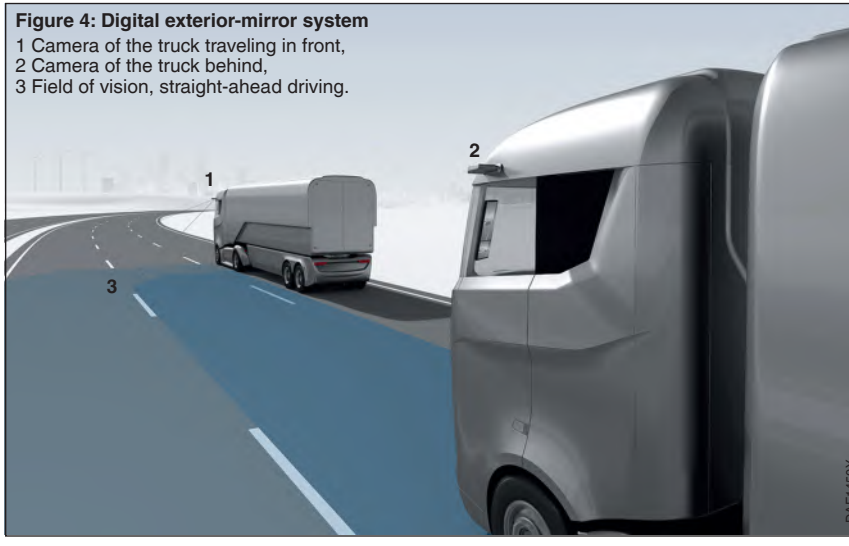
Digital exterior-mirror system

Large side mirrors on commercial vehicles are an extremely important safety feature and are essential for ensuring the driver has a good rear view. But, due to their size, they restrict forward visibility and also generate high aerodynamic drag. The digital exterior-mirror system replaces the large side mirrors with two monitors integrated in the driver's cab and two exterior cameras that are mounted above the driver's cab (Figure 4). The aerodynamic design reduces aerodynamic drag and with it fuel consumption. At the same time, the system increases safety by significantly improving the driver's all-around visibility and reducing the blind spot.

Replacing the two exterior side mirrors provides additional advantages with the innovative technology. The technology of the digital exterior-mirror system furthermore makes it possible to provide a situation-specific display:

– on expressways and highways the driver can see further behind the vehicle,
– in towns and cities the viewing angle increases to expand the driver's field of view,
– at night the enhanced contrast improves visibility.

Figure 4: Digital exterior-mirror system
1 Camera of the truck traveling in front,
2 Camera of the truck behind,
3 Field of vision, straight-ahead driving.



Road-sign recognition

Features
Another function that makes use of the information from the video sensor is road-sign recognition. This system is capable of recognizing and interpreting road signs (e.g., speed limit or no-overtaking signs). A video camera takes over 20 images per second and transmits these in the form of a video signal to the image-processing computer. While the vehicle is moving, the system searches for objects which on the basis of their external shape could be road signs (object detection). When such an object is found, it is tracked until it is close enough to be picked up by the video camera as a specific object (object classification). These can be road signs which show for example speed limits, the start and end of stretches where passing (overtaking) is prohibited, and also no entry. Stop, right-of-way, and road-works signs are also recognized. Road-sign recognition also classifies relevant supplementary signs, such as time limits, signs applicable only to specific vehicle types, and turn arrows. The system reliably detects road signs, whether they are sited on actual signs, variable message systems or gantries. Road-sign recognition presupposes that the sample designs to be recognized have previously been taught into the image-processing computer.

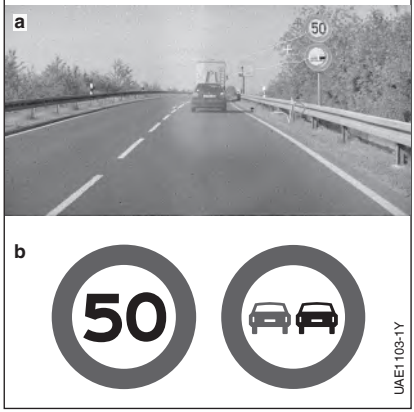
When the system detects a relevant road sign, it shows this as an icon in the instrument cluster's graphic display (Figure 5). If the driver for example fails to observe the detected speed limit, the system can issue an additional acoustic or haptic warning. The driver can also be warned against passing (overtaking) on a stretch where such maneuvers are prohibited and when stop and no-entry signs have been overlooked.

Reliable road-sign recognition is now possible at speeds of up to 160 km/h and also in conditions of rain and road spray.

Application In assistance functions
Recognized road signs are increasingly being used in a variety of assistance functions. For example, navigation maps can be updated with recognized road signs and the adaptive cruise control system can in a further expansion stage maintain the vehicle's speed at the identified speed limit.

When a pedestrian is identified as an object class, the function can become active for active pedestrian impact protection if there is an imminent danger of a collision (see Active pedestrian protection).

Figure 5: Road-sign recognition
a) Road image,
b) Icons in instrument-cluster graphic display.



Driver-drowsiness detection

Problem
Fatigue and microsleep at the wheel are often the cause of serious accidents. However, the initial signs of fatigue can be detected before a critical situation arises. The steering behavior can provide an indication of the onset of driver drowsiness.

Operating principle
Using information from the steering-angle sensor, the drowsiness-detection algorithm analyzes the driver's steering behavior and recognizes changes that occur over the course of long trips and in response to driver fatigue. Typical signs of waning concentration are phases during which the driver is barely steering, combined with slight, yet quick and abrupt steering interventions to keep the car on track. Already small steering deviations that are barely perceptible to the driver are typical signs of waning concentration. Based on the frequency of these steering corrections and other parameters, among them the length of the trip, blinking behavior, and the time of day, the function calculates a fatigue index. If this index exceeds a certain value, an icon such as a flashing coffee cup in the instrument cluster alerts the driver to take a break.

This drowsiness-detection system has already successfully been on the market for a few years. Because it is purely a software solution in combination with the steering-angle sensor, the function can be integrated with minimal effort and outlay in existing vehicle platforms. The function helps to fulfill the upcoming legal requirements.

A more sophisticated realization of driver-drowsiness detection is based on camera-based interior monitoring.

Rear cross-traffic warning

Features
Backing out of a parking space (Figure 7) can often be a challenge, especially when the driver's view of the traffic passing behind their own vehicle is obstructed. Rear cross-traffic warning makes this reversing parking maneuver easier by warning the driver of crossing vehicles. Rear cross-traffic warning can detect crossing vehicles up to 50 m away.

Operating principle of rear cross-traffic warning
The system utilizes two corner radar sensors in the rear of the vehicle, Figure 6 shows the detection range. They measure and interpret the distance, speed, and anticipated driving path of detected vehicles in the cross-traffic.

When the function detects other vehicles crossing from the left or right at a distance of up to 50 meters, it warns the driver acoustically or visually of an imminent danger of collision.

Figure 6: Detection range of rear cross-traffic warning



System benefits for the car driver
– Rear cross-traffic warning warns the driver of crossing vehicles when backing out of parking spaces,
– it facilitates safe and stress-free reversing out of parking spaces, even with limited visibility.

Turn collision warning for trucks

Features
Driving heavy commercial vehicles can be a challenge, especially at intersections. Cyclists and pedestrians are particularly at risk in such situations. Turn collision warning promptly detects pedestrians and cyclists in the commercial vehicle's blind spot (Figure 8), warns the driver acoustically or visually of potential dangers before the vehicle turns, and thereby helps to avoid accidents. The side radar sensors can not only facilitate this function, but can also more than satisfy the legal requirements.



Figure 7: Backing out of a parking space with cross-traffic



Figure 8: Turn collision warning for heavy trucks

Operating principle of turn collision warning

Radar sensors installed on the side of the vehicle promptly detect pedestrians and cyclists on the co-driver's side. These sensors cover the entire area alongside the vehicle and trailer, from the immediate area in front of the vehicle reaching into the intersection to just behind the end of the trailer. If the side-mounted radar sensors detect other road users in these areas, the system outputs a visual signal to notify the driver of their presence. In this situation, the vehicle's predicted route is tracked against the road users detected.

If a person is situated in the critical area alongside the vehicle and the driver initiates a turning maneuver, the driver receives a further visual or acoustic warning. This enables the driver to react in good time to avoid a collision with a pedestrian or cyclist. The system helps to increase road safety and prevent accidents between heavy commercial vehicles and more vulnerable road users. With these side-mounted radar sensors, the legal requirements for protecting pedestrians and cyclists during turning maneuvers contained in regulation UN ECE R151 [1] for “blind-spot information systems” can be more than satisfied.

System benefits for the vehicle driver

– Turn collision warning increases road safety by displaying warnings of hard-to-see road users or persons obscured by the vehicle,
– it assists the driver with acoustic or visual warning signals.

Moving-off information system for heavy commercial vehicles

Features

Maintaining a clear view in heavy commercial vehicles can be a challenge, especially in urban traffic. Cyclists and pedestrians are particularly at risk in such situations. The moving-off information system helps to avoid collisions. It detects objects and people within close range of the front of the vehicle (Figure 9), issues the driver with acoustic or haptic warnings of potential dangers, and therefore helps to prevent accidents. The side radar sensors can not only facilitate this function, but can also more than satisfy the legal requirements.

Operating principle of the moving-off information system

The moving-off information system is a cornerstone for integrative, near-range safety functions and helps to avoid collisions with other road users. It provides the driver with an assistance system that detects people and obstacles close to the front of the vehicle in situations involving moving off and low-speed maneuvers. Side-mounted radar sensors at the front of the cab constantly monitor the vehicle's immediate surrounding in such situations.

If the sensors detect objects or road users and recognize a potential risk of collision, the system promptly provides the driver with an acoustic or haptic warning. This enables the driver to react swiftly and brake in good time. The visual or acoustic warning ceases when there is no further risk of collision. With these side-mounted radar sensors, the legal requirements of the General Safety Regulation (GSR) can be more than satisfied.



Figure 9: Moving-off information system for heavy trucks

System benefits for the vehicle driver

– The moving-off information system increases road safety by displaying warnings of hard-to-see road objects or objects obscured by the vehicle,
– it assists the driver with acoustic or haptic warning signals,
– it increases safety for all road users.

Wrong-way driver warning

Motivation

In Europe there are more than 4,000 accidents every year caused by wrong-way drivers on expressways/highways. This amounts to around 5 % of all the fatal accidents on expressways/highways. The aim of cloud-based wrong-way driver warning is to avoid the often serious accidents involving wrong-way drivers on expressways/highways. This function thus makes an important contribution to increasing road safety.

Technical implementation

In order to reach a very large number of road users as quickly as possible with the service, an approach was chosen which called for solely connectivity and GPS (Global Positioning System). The function can be integrated both in mobile end devices such as smartphones and directly in the vehicle's central communication unit (head unit). To this end, all that needs to be integrated is a “Software Development Kit” (software comprising tools and libraries), which contains the coordinates of all the relevant highway entrance ramps and the necessary algorithms for performing the steps described in the following.

Functioning principle

As soon as a networked vehicle reaches the area of a high entrance ramp, the end device starts to send the position and acceleration data in one-second intervals to the back end. Here the actual direction of travel is compared with the permitted direction of travel. The data required for this purpose are available in the cloud in the form of a map with direction-of-travel details. A warning is sent if the two sets of information clash. This warning is sent to all the connected vehicles. Only then is a check made in the end device as to whether the individual vehicle is affected by the warning and this warning must be displayed to the user accordingly. It is up to the customer to organize the nature and the content of the warning. Bosch in its proposal is orientated to the recommended actions of the German motoring association ADAC.

Performance

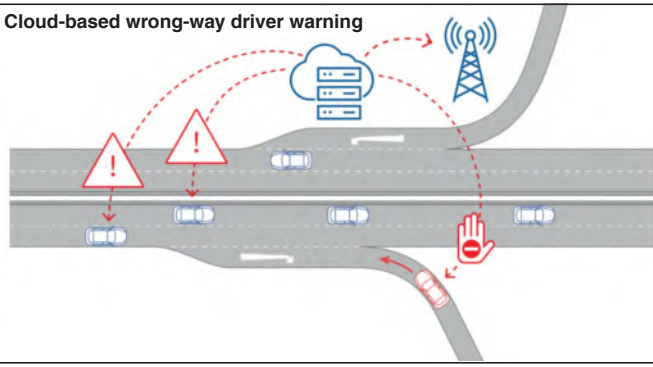
In this way, it is possible to warn not only the wrong-way driver themselves but also quite specifically those road users who are driving towards the wrong-way driver.

With a warning time of under 10 seconds with a “true positive rate” of 93 % (actual wrong-way drivers detected) and a “false positive rate” of just 0.0004 % (incorrect warnings are issued in only four out of every one million cases) this solution is more efficient than other approaches to preventing wrong-way driving incidents.

References

[1] UN ECE R151: Blind Spot Information System for the Detection of Bicycles.

Figure 1: Cloud-based wrong-way driver warning



Lane assistance

Drifting from the lane unintentionally frequently leads to serious accidents. It is caused for the most part by the driver being distracted or tired (e.g., microsleep). Lane departure warning (LDW) is designed to help avoid such accidents by detecting the lane markings ahead and warning the driver if there is a danger of a lane marking being crossed without the direction-indicator lamp (turn signal) having been activated. The lane-keeping support system (LKS) also actively intervenes in vehicle guidance (Figure 1).

Assistance systems that afford the driver further lateral-guidance support on expressways/highways with roadworks (roadworks assistant) or in urban traffic (bottleneck/constriction assistant) are currently in development.

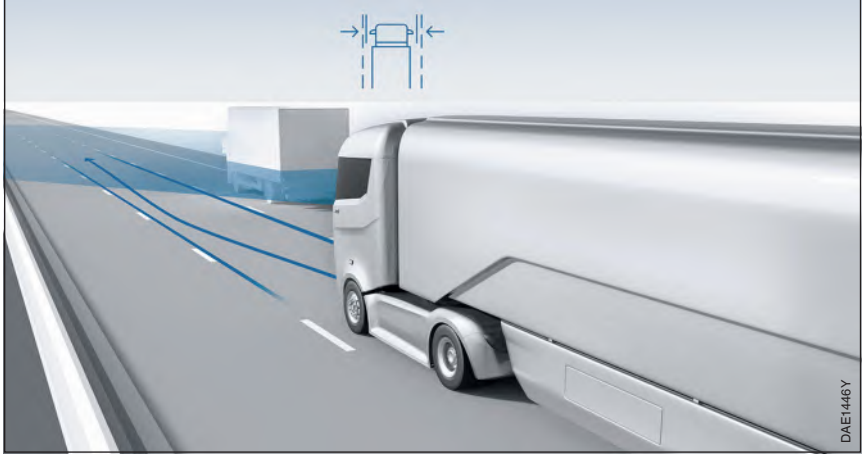
Lane departure warning

Lane detection

Lane-departure warning systems use video cameras to detect the lanes ahead of the vehicle and can have both mono and stereo video designs. A stereo video sensor offers advantages when the road surface exhibits more pronounced gradients, as is typical for example of many country roads. Here the stereo video sensor helps to estimate the surface profile. Knowledge of the surface profile simplifies in the next stage the detection of lane markings and extends the detection range. There are no advantages for lane detection on expressways/highways as here only slight changes in the surface profile are permitted by statutory requirements. In good weather conditions and with good lane markings, the range is in the region of up to 100 m.

Figure 2 shows the principle of video-based lane detection. The image-processing system searches for lane markings by analyzing differences in contrast between the road surface and the lane markings. Figure 2a shows the camera image with search lines, Figure 2b shows a detailed view. The crosses in Figure 2a

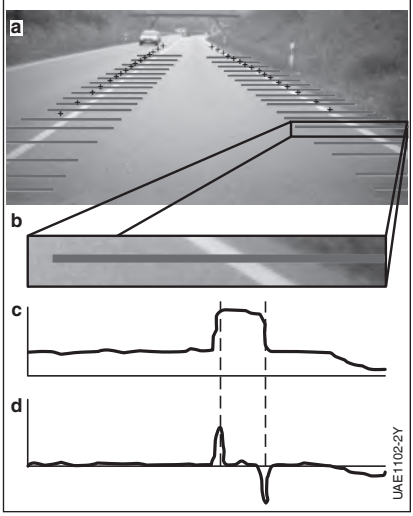
Figure 1: Lane-keeping support for heavy trucks



mark the course of the lane which have been calculated by the image-processing computer. In order to detect a line, the luminance signal (brightness) inside the search line is analyzed (Figure 2c). The limits of the lane markings are detected by high-pass filtering (Figure 2d). The lane covered by the system vehicle can be detected on the basis of the limiting lane markings. A brief absence of lane markings can be compensated for in the currently available systems. No lane assistance can be offered if the markings are absent over wide areas.

The lane-detection algorithm of the camera detects and classifies all the common variants of lane markings up to a distance of around 60 meters (in good visibility conditions round 100 meters), regardless of whether they are continuous, dashed, white, yellow, red or blue. Even “Botts‘ Dots” (round, non-reflective, raised lane markers) can be detected by the camera.

Figure 2: Principle of lane detection
a) Camera image with search lines,
b) Detailed view with search line,
c) Luminance signal (high level for light lane marking),
d) Edge information by high-pass filtering of luminance signal (peaks at light/dark transitions).



Research is currently being conducted into systems which are intended to detect the roadsides even without markings on the basis of the surface structure and color. The attributes of the last road driven on are also utilized as models. When such systems are market-ready in a few years, lane-assistance systems will also be available on unmarked roads.

Driver warning
A warning for the driver can be derived on the basis of the detected lanes in the event of the vehicle drifting out of lane. Different forms of warning are conceivable here. The first systems used an acoustic warning from the vehicle speaker in the form of a warning sound or “rumblestrip rattling”; a stereo sound can additionally convey a direction here. However, the systems that have become the most widely used since then feature warning the driver via the haptic sensory channel by vibrating the steering wheel or subjecting the steering to a slight counter-torque. The advantage of issuing the warning through the steering wheel lies in the driver directly associating the danger with the movements on the steering wheel.

Regulations and standards
Together with the mandatory introduction of the automatic emergency-braking system for heavy commercial vehicles (8 t and upwards) the Lane Departure Warning System (LDWS) has been prescribed for this vehicle class in the EU (European Union LDWS Regulation (EU) 351/2012 [1]). Here too the introduction was conducted incrementally from 1 November 2013 initially for vehicles to be newly homologated and from 1 November 2015 for all new vehicles brought onto the roads.

Lane-keeping support

Supporting interventions
Lane-keeping support is a functional enhancement of lane departure warning in that it not only warns the driver against drifting unintentionally from the lane, but also actively supports him in maintaining course. If the vehicle drifts too far from the lane, lane-keeping support makes a course correction, either by means of an active steering intervention or an asymmetrical braking intervention. In this way, the system supports the driver in keeping his vehicle on course and at the same time helps to increase road safety by avoiding accidents caused by drifting unintentionally from the lane.

However, lane-keeping support itself cannot assume full lateral guidance of a vehicle, but is instead intended to support the driver in maintaining course. The system does not relieve the driver of his responsibility to follow and react to the traffic conditions attentively and to drive the vehicle consciously and deliberately. A system intervention can be overridden by the driver at any time.

To prevent misuse, the lane-keeping support system is fitted with a device which monitors whether the driver has both hands on the steering wheel and is actively steering the vehicle. If the driver is driving with no hands, an instruction is promptly issued to the driver to take control of the vehicle and the assistance afforded by lane-keeping support is terminated with an accompanying visual or acoustic warning.

Applications
Lane-keeping support is designed primarily for use on expressways/highways and well developed highways. Like lane departure warning, lane-keeping support is dependent on detecting lanes ahead of the vehicle. Short-term failures in lane detection, caused for example by hidden or faded lane markings, are compensated for by the system. However, lane-keeping support is deactivated if no lane information is received for an extended period of time.

Lane centering assist for heavy trucks

Application
Tiredness and distraction are the most frequent causes of a driver unintentionally drifting over lane markings, especially on long and monotonous journeys. Event just a fleeting moment of inattention can lead a driver to depart from their lane. Lane centering assist prevents this by actively keeping the vehicle in the center of its lane (Figure 3), bringing both driver and load to their destination safely and comfortably.

Lane-keeping support actively steers the vehicle in the center of its lane.

Operating principle of lane centering assist
Lane centering assist uses a video camera to detect lane boundaries – including road markings or structural boundaries – alongside the vehicle and compare these with the position of the vehicle in the lane. Then the system aligns the vehicle via the electrohydraulic steering system to keep it on track, making active interventions to ensure that it stays in the center of its lane.

The function can be activated while driving and – if necessary – can be overridden by the driver at any time. In this way, the driver remains responsible for the vehicle.

The smart combination of components makes driving heavy commercial vehicles safer and brings a noticeable improvement in driving comfort. Damage to the load caused by jerky steering maneuvers is avoided.

Figure 3: Lane centering assist for heavy trucks



Emergency lane-keeping support for heavy trucks

Application
Long, monotonous stretches of road can cause drivers to become tired and inattentive. On occasion, this can make them unintentionally drift over the lane markings. Things get particularly dangerous when unprotected, steep embankments or verges line the edge of the road. If the driver starts to drift away from the drivable road surface and fails to react promptly, emergency lane-keeping support actively steers the vehicle back into the correct lane.

Operating principle of emergency lane-keeping support
Emergency lane-keeping support uses a camera to detect the edge of the road, i.e., the boundary of the surface that can be safely driven on, and compares this with the vehicle's position. The function intervenes if it determines that the vehicle is failing to maintain a defined minimum distance to the boundary of the road surface.

Emergency lane-keeping support makes a noticeable steering correction in the opposite direction with the electrohydraulic steering system to keep the vehicle on track and thereby avoid the serious consequences of the vehicle departing the road. It increases safety through early correction of driving mistakes.

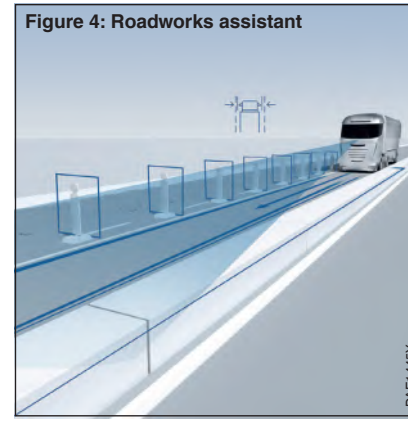
Roadworks assistant

Applications
When roadworks are taking place on the roads, there is often very little space between the vehicles in the adjacent lane and the crash barriers. Roadworks assistant and bottleneck/constriction assistant are further developments of lane-keeping support. In addition to lane markings, these systems also detect raised objects and obstacles and take them into consideration in lateral guidance (Figure 4). The help drivers maintain a lateral safety distance to vehicles in the adjacent lane and safety curbs or crash barriers when driving in a narrow lane.

The roadworks assistant supports the driver in negotiating roadworks on expressways/highways and thus extends the lane-keeping support system to include this special scenario.

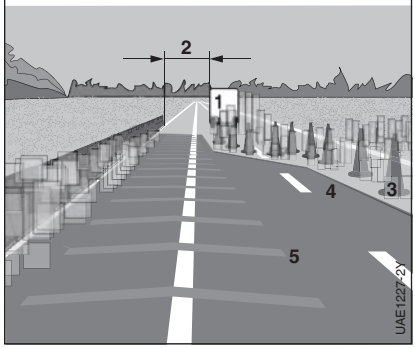
Types of support
The type and the extent of support provided can vary greatly. The system can run through all the stages of a cascade of intervention, from the visual, acoustic or haptic warning right up to interventions in vehicle guidance. An intervention in vehicle guidance can involve interventions both in the lateral vehicle dynamics (steering torque or asymmetrical braking intervention) and in the longitudinal dynamics (deceleration). Automatic deceleration can help to keep the vehicle in lane if it drives too fast into a roadworks chicane.

Figure 4: Roadworks assistant



Detection of lane restrictions
Because roadworks are more complex in their structure than the typical expressway/highway environment, the roadworks assistant detects roadworks-specific lane restrictions (traffic beacons, cones, safety barriers) in addition to the regular lane markings. The sensor functions are typically performed by a front-end stereo video camera and side-mounted ultrasonic sensors. These sensors enable the system to distinguish between static objects (e.g., roadside structures) and dynamic objects (e.g., vehicles). In this way, the system can take the necessary collision-avoidance action in the event of lane constrictions at roadworks caused by traffic in neighboring lanes. This in turn affords the driver support to suit the situation as he or she drives through expressway/highway roadworks and as a result offers the best support possible for maintaining a sufficient safety distance.

Figure 5: Bottleneck/constriction assistant – visualization of calculation results
1 Inflatable obstacle (balloon car),
2 Bottleneck/constriction,
3 Collision-relevant objects (e.g., cones, safety curbs),
4 Available road width,
5 Predicted corridor.



Bottleneck/constriction assistant

Applications
The bottleneck/constriction assistant extends the roadworks assistant with regard to the typical speed range in urban areas (0 to 60 km/h) and with regard to the supported radii of bends. Collision-avoiding steering assistance is offered in urban bottlenecks/constrictions accordingly.

Functionally negotiable bottlenecks/constrictions can also be caused by objects traveling in the same direction or by oncoming objects. The bottleneck/constriction assistant provides automatic steering interventions so that a sufficient safety distance from side obstacles is maintained when passing through tight spaces (bottlenecks, constrictions). It also ensures that an early warning is issued to the driver and an automatic brake intervention is effected if there is not enough space for the vehicle to be able to pass.

As with lane-keeping support, the intensity of the steering and brake interventions is restricted. The interventions can be easily overridden by the driver so that potentially motiveless interventions remain fully non-critical and thus without consequences.

The feasibility of the functional idea was demonstrated by the implementation of an in-vehicle prototype as part of the publicly funded UR:BAN project concluded in 2016 [2]. The prototype's system response in a typical urban scenario is shown in a video in [3]. The bottleneck/constriction assistant is currently in the advance-development stage. The date of a series launch is still open.

Environment detection
A functional requirements analysis reveals that a forward-mounted stereo video camera and alternatively side-mounted ultrasonic sensors or monocular video sensors must be used. The stationary environment and moving objects (those traveling in the same direction and oncoming objects) must be represented within the framework of environment modeling. An occupancy grid presents itself for the stationary environment. An occupancy grid is a metric grid of the vehicle's environment.

Typical cell sizes range between 10 cm and 20 cm. If an object exists in the system vehicle's surrounding environment, the affected grid cells are marked as occupied. Dynamic objects are pictured by way of object models with various function-related attributes (position, width, length, speed, and acceleration).

Evaluation of the traffic situation

The situation analysis must make an assessment of the collision risk in accordance with the planned trajectory of the vehicle and the moving objects, and incorporating the stationary environment. Criticality measurements which are used in the function module to detect critical situations where action is required are typically calculated for this purpose. A typical criticality measurement of this type is the braking-time reserve (also called "time to brake"), which specifies how much time the driver still has to avoid, through a self-initiated brake intervention, a collision for example in an unpassable bottleneck/constriction. If the driver is no longer able to react (i.e., the braking-time reserve is zero or negative), a collision-mitigating system response is authorized.

Response model

The function module checks, aside from the criticality measurements mentioned, that the system limits are observed (e.g., whether the maximum supported bend is exceeded) and activates where necessary steering and brake assistance or issues an early warning to the driver.

Figure 5 visualizes some central calculation results by the bottleneck/constriction assistant in a test scenario with a passable bottleneck/constriction while using typical test equipment (safety curbs, cones, balloon car). The rectangles represent detected collision-relevant objects. The gray area (in the graphic on the left of the pictured boundary line) codes the available road width in the bottleneck/constriction. The stripes on the level show the predicted corridor for the vehicle.

References

[1] Commission Regulation (EU) No. 351/2012 of 23 April 2012 implementing Regulation (EC) No. 661/2009 of the European Parliament and of the Council as regards type-approval requirements for the installation of lane departure warning systems in motor vehicles.
[2] www.urban-online.org
[3] UR:BAN research initiative (2015) https://youtu.be/eyvh43Jq5yA?t=37s

Lane-change assistant

The number of passenger-car accidents caused by inappropriate lane changing was in the past years regularly in the range of between 6 % and 8 % [1]. Here, the incidence of accidents is mainly centered on highways and expressways/highways. Drivers initiate a lane change and fail to notice other road users either next to or behind their vehicle. In the first case, the other road users are outside the field of vision of the side mirrors (blind spot, see Figures 1 and 2). On expressways/highways the colliding vehicles are typically driving at similar speeds, i.e., are traveling parallel or slowly passing (overtaking). In the second case, a lane change is made in front of a vehicle that is approaching from the rear at high speed. A typical outcome here is a rear-end collision.

Subfunctions which can address the two types of accident mentioned are described in the following. The functions described are available for trucks and motorcycles as well as for passenger cars. The principle of operation is the same, regardless of the vehicle type.

Blind-spot assistant

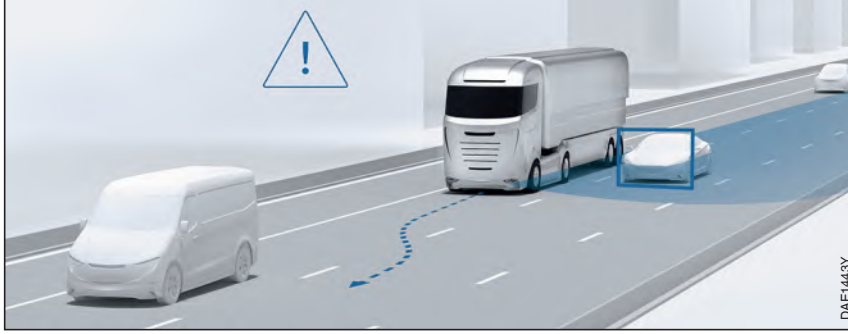
The blind-spot assistant helps to avoid the first type of accident mentioned.

Environment detection

There are various sensor concepts, differing in terms of costs and performance. In order to avoid this type of accident, only the area directly next to or just behind the system vehicle is relevant. What all the sensor concepts have in common is that the range to the rear extends only a short distance beyond the rear end of the vehicle.

There are systems that utilize the ultrasonic sensors (also used for the parking aid) at the rear corners of the vehicle. As a result, small objects (e.g., scooters) which are situated directly next to the system vehicle cannot be measured, since these are outside the ultrasonic sensor's field of vision. Another solution available on the market utilizes mono cameras in the side mirrors. In a third variation, short-range radars to the sides are used. In the last two cases mentioned even smaller objects at the sides of the vehicle can be detected by sensors. It is to be noted that the sensors mentioned can be used again by the vehicle manufacturers in a value-added approach for a variety of functions (e.g., parking aid, turning assistance, integral safety).

Figure 1: Blind-spot detection in a truck – the entire area alongside the trailer is detected

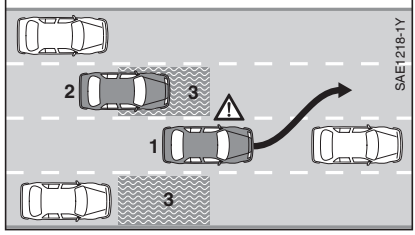


Sensor-related challenges arise in driving maneuvers that involve high transverse dynamics (turning operations). Here the manufacturers typically set system limits (e.g., thresholds for lateral acceleration, yaw rate, radius of curve which when exceeded deactivate the system) in such a way that incorrect object measurements and consequently false system warnings are avoided.

Response model
The following warning cascade will run if the sensor system detects a vehicle in the neighboring lane.

In a first stage, only slightly perceptible driver information is provided in the form of a side-specific activation of visual symbols or light elements in the side mirrors. If the driver additionally signals a lane-change request (e.g., by activating the direction-indicator lamp or on the basis of the measured driver steering torque on the steering wheel), the visual display can flash, followed by a supplementary acoustic warning or even a slight intervention in lateral guidance. The latter steering intervention can actively prevent or impede the lane change. Automatic steering intervention can be easily overridden and is comparable in terms of intervention intensity with that of lane-keeping support.

Figure 2: Blind-spot detection in a passenger car
1 Vehicle pulling out,
2 Vehicle in the detection range of vehicle 1,
3 Detection range of vehicle 1.



Lane-change assistant

As described above, the blind-spot assistant, on account of the limited sensor field of vision to the rear, cannot react to vehicles approaching at high speed from the rear (with the typical outcome of a rear-end collision). Here, the function of the lane-change assistant supplements the range of functions of the blind-spot assistant, i.e., simultaneously monitors the space to the sides of the vehicle as well.

Environment detection
In contrast to the blind-spot assistant, exclusively mid- or long-range radars are used here; these are mounted on a passenger car at the rear corners below the bumper. If a truck is fitted with a lane-change, this is typically positioned under the side mirror.

In particular for passenger cars the radar-antenna radiation pattern is typically set out in such a way that the radar sensor with a smaller radar lobe of low range also covers the area next to the vehicle. The radar sensors use attain a range of up to 100 m to the rear.

Response model
Unlike with the blind-spot assistant, an object prediction is now conducted on the basis of the object list created in environment detection. The system therefore measures whether at a given system-vehicle speed the object approaching from the rear would enter into a critical situation, i.e., would have to brake sharply, if the system vehicle changes lane. The braking-time reserve (time to brake, see Bottleneck/constriction assistant) of the approaching object is determined as the criticality measurement for this purpose. If a lower threshold is undershot (typically <2 to 3 seconds), the lane change is assessed as critical and the escalation cascade passes, as with the blind-spot assistant, into the second stage. The actual response model now corresponds to that of the blind-spot assistant.

Evasive steering support

Features
Evasive steering support actively supports the driver during a critical evasive maneuver by means of a supporting steering torque. The helps the driver to circumnavigate an obstacle in order to avoid an accident. The maximum steering angle is reached faster with the assistance of evasive steering support.

Evasive steering support becomes active when the system detects an impending collision and the driver initiates an evasive maneuver. Once the driver begins steering, the system assists the driver with torque interventions in the power steering in order to guide the vehicle around the obstacle (Figure 3).

Using the data from the video camera and radar sensor, the system instantaneously computes a suitable path, taking into account distance, width and offset of the vehicle traveling in front. The collected data from the radar and the camera are merged in an ECU for a better and more robust understanding of the scene.

System benefits for the car driver
– Evasive steering support reduces the risk of collisions with objects and thus associated injuries,
– it supports the driver in maintaining a safe evasive path,
– it maintains driving stability.

Further developments

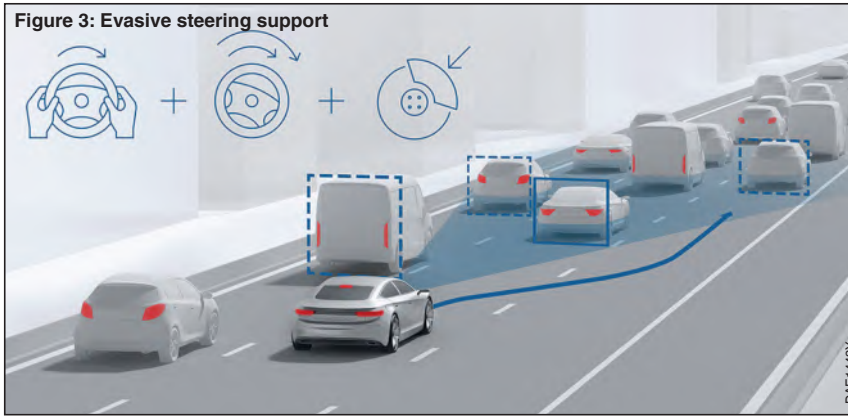
Automation of the lane-change operation – particularly on expressways/highways – is conceivable with the above-mentioned lane-change assistant. Such systems are currently in development.

Here the system indicates to the driver the possibility of a safe lane-change maneuver provided the system vehicle is following another slower vehicle. The driver enables the maneuver and can gently place his or her hands on the steering wheel. The system performs the lane change automatically. In combination with Adaptive Cruise Control (ACC) the system would speed up the vehicle and then after the overtaking operation is completed get the vehicle back in lane in front of the passed (overtaken) vehicle.

Since the radar sensor monitors the traffic behind the vehicle, such a maneuver can only be initiated if there is no risk of a side collision. This significantly increases overall safety. Most of the accidents that currently occur due to driver error when changing lanes could thereby be avoided.

References
[1] GIDAS (German In-Depth Accident Study) 2014, http://www.gidas.org

Figure 3: Evasive steering support



Emergency-braking systems

Emergency-braking systems in longitudinal traffic

Emergency brake assist and automatic emergency braking

A significant number of road accidents are the result of collisions between vehicles driving in the same direction, i.e., rear-end collisions. Driver-assistance systems which can avoid such accidents or at least mitigate their consequences have been in series production for some years now. A simple feature is the brake assistant (see Brake-assistance functions). Systems that use sensor technology to observe and interpret the vehicle's surroundings have a particularly high potential [1]. This enables such a system to make the necessary anticipatory intervention in vehicle guidance in order for example to avoid an accident with emergency braking.

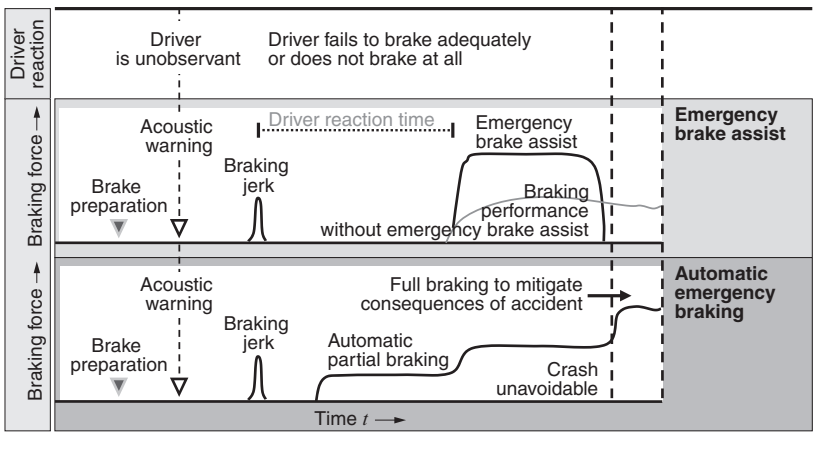
Environment detection

To be able to avoid rear-end collisions, it is necessary to observe the vehicle traveling in front and its movement. If the system detects that a collision is imminent, it can take appropriate action, for example by issuing a warning to the driver or by performing emergency braking.

Various sensors are suitable for observing the vehicle environment. Their range and sensing range constitute an important function-relevant feature. The required look-ahead distance of the sensors is derived above all from the speed range in which the function is to be active. Radar sensors are used predominantly used for long ranges which are required at high driving speeds. Mono and stereo video sensors typically achieve medium look-ahead distances. Lidar sensors can be used for both medium distances and close range.

Several sensors are frequently combined for a automatic emergency-braking system. This offers two advantages: Firstly, the strengths of the individual measurement concepts complement each other, providing for a more accurate measurement of vehicles traveling in front. Secondly, system reliability is enhanced in that possible faults in individual sensors can be detected. In particular, unjustified system responses can be avoided. It is also possible, albeit with a loss in per-

Figure 1: Braking-response model of Emergency Brake Assist and Automatic Emergency Braking for an unobservant driver who fails to brake adequately or does not brake at all



formance, to realize emergency-braking systems which are based merely on the data from a single sensor.

Attributes of objects traveling in front such as position, speed, acceleration, and object size and type are assessed on the basis of the environment-sensor data and taking into account the motion of the driver's own vehicle. These object attributes are then used to evaluate the traffic situation with regard to its criticality.

Evaluation of the traffic situation
Different measures have become established that describe how critical the current traffic situation is, i.e., how probable it is that a collision with a vehicle traveling in front is imminent (criticality measurements). Assumptions regarding the future motion both of the vehicle traveling in front and of the driver's own vehicle are made for this purpose. This is possible since only very short prediction horizons are required for emergency-braking systems. It can be assumed for instance that the driver of the actual vehicle will still try by means of an avoidance maneuver to avoid the collision. Assuming maximum transverse dynamics for this avoidance maneuver, the last possible moment when the driver can still avoid the collision by evasive maneuver can be calculated. The time remaining until this moment is called the "time to steer" (TTS). Similarly, a "time to brake" (TTB) can be calculated, i.e., the time remaining to the driver to still avoid the imminent collision by means of a braking maneuver. The shorter these times, the more critical the traffic situation. There are many more criticality measurements, e.g., the probability of a collision in the future "time to collision" (TTC). To be able to calculate these, it is necessary to make assumptions about probability distributions for the future motion of the vehicles involved.

Response model
Different system responses can be initiated if a collision is imminent and this is detected when the above-mentioned criticality measurements are undershot. The time sequence of an exemplary system response is illustrated in Figure 1. The following system-response stages exist in standard systems:

Brake preparation
If the probability of an imminent collision increases, the brake system can be prepared by driving-dynamics control (Electronic Stability Control, ESC) in such a way (prefill) that any subsequent braking (by the driver or by the system) can be effected more rapidly.

Warning
The driver can be warned in the event of a heightened collision risk. A warning is suitable to force the driver into a tried-and-tested response, especially if he has not yet recognized the danger. An accident can thus be avoided by the driver himself if the advance-warning time is long enough.

However, when configuring a warning function, it is important not to set the advance-warning time at too long; in such a case the warning would frequently be issued already in easy-to-control, non-critical situations and would thus be perceived to be unjustified.

Braking jerk
The braking jerk constitutes a further warning escalation stage. Here the brake system is automatically activated for a brief period and then released again immediately. This has two advantages: Firstly, the driver is warned by haptics of the danger; secondly, the driving speed is already slightly reduced.

Partial braking
Partial braking works along similar lines to the braking jerk. Here a slight deceleration is automatically set, but unlike the braking jerk is not released again immediately.

Brake assistant
Frequently the driver will react to the critical situation by initiating a braking operation that is too weak to avoid the accident. In this event the system can boost the braking force requested by the driver so as to bring the vehicle to a stop in good time before the obstacle and avoid the accident.

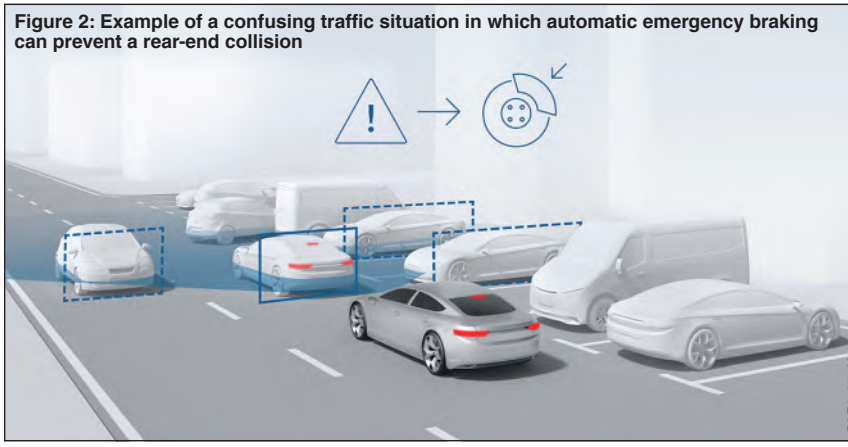
Full braking
If the system responses described above are not sufficient or the driver fails to react appropriately to avoid the imminent collision, the above-described criticality measurements do not decrease sufficiently – the situation therefore remains critical. In this case, automatic full braking is initiated at the last moment. A possible criterion for activating emergency braking is for example a “time to brake” which is virtually zero or negative. Automatic full braking will frequently prevent an accident entirely, but at least will greatly reduce the impact speed.

Regulations and standards
Emergency-braking systems for passenger cars
Since 2014 automatic emergency braking when driving in the same direction has been part of the Euro NCAP assessment (European New Car Assessment Program). Emergency-braking systems are assessed by reference to standardized collision scenarios which are recreated on a test track. Here, both the avoidance of collisions by an emergency-braking system and a reduction of the impact speed are positively assessed. The effectiveness of the emergency-braking system installed in each case finally affects the number of NCAP stars that a vehicle can achieve.

Emergency-braking systems for commercial vehicles
An automatic emergency-braking system (AEBS, Advanced Emergency Braking Systems) has since been prescribed in the EU for heavy commercial vehicles weighing 8 t and upwards (European Union AEBS Regulation (EU) 562/2015, [2]). The introduction was conducted incrementally from 1 November 2013 initially for vehicles to be newly homologated and from 1 November 2015 for all new vehicles brought onto the roads. In a further step the requirements with regard to reducing the speed in respect of moving and stationary targets were tightened from 2016.

Available systems
Various systems are currently available which can significantly reduce speed in critical situations and frequently avoid the collision entirely. However, an automatic emergency-braking system can involve the additional risk of a rear-end impact and in the event of unjustified deployments give rise to acceptance problems. Suitable measures must therefore be taken within the framework of system development to ensure that no unauthorized emergency braking occurs (see “Test and hedging”).

Figure 2: Example of a confusing traffic situation in which automatic emergency braking can prevent a rear-end collision



Statutory requirements
With the front radar sensor and optionally the multifunction camera the legal requirements of UN-ECE R131 [3] regarding the Advanced Emergency Braking System (AEBS) and a large number of regional legal requirements can be satisfied.

Maneuver emergency braking

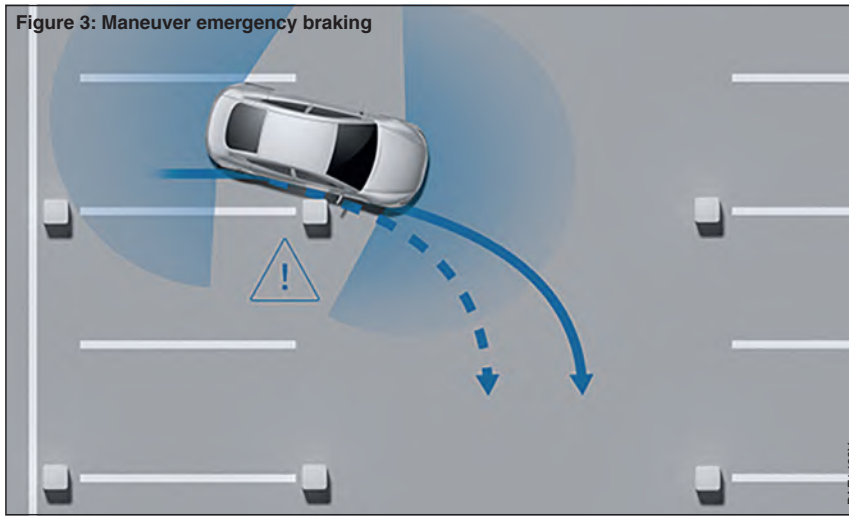
When maneuvering in tight and complex driving situations and when attempting to maneuver the car into a parking space, maneuver emergency braking prevents the vehicle from bumping into posts, pillars, and neighboring vehicles.

Operating principle of maneuver emergency braking
The assist function constantly monitors the vehicle's surroundings by means of ultrasonic sensors. Using steering angle and driving speed, the system calculates the car's intended path and identifies any obstacles that could get in the way while maneuvering into the parking space (Figure 3).

In the event of an imminent collision, the assist function automatically applies the brakes promptly until the car comes to a stop, thereby preventing dents or scratches.

System benefits for the car driver
– Maneuver emergency braking supports the driver when maneuvering in tight driving situations,
– it helps to prevent accidents when parking and maneuvering and thus avoid damage to the sides of the vehicle.

Figure 3: Maneuver emergency braking



Automatic emergency braking for unprotected road users

Active pedestrian protection

Motivation

The protection of vulnerable road users, particularly pedestrians, has increasingly taken center stage in recent years. Initially this increased focus resulted in the improvement of passive safety systems for pedestrian protection (e.g., specific design of the vehicle front end to mitigate the consequences of collisions with pedestrians). Now, increasingly, active systems – i.e., systems which clearly act against potential collision – are also going into series production.

Active pedestrian-protection systems react similarly to emergency-braking systems for vehicles traveling in the same direction. However, the environment-sensing requirements are different; this can be identified in an analysis of the frequently occurring types of accident.

Causes of accidents

The majority of pedestrian accidents involve persons who cross the road while the vehicle is driving straight ahead. The pedestrian steps out from the side in front of the moving car (Figure 4), often from a concealed position (i.e., the pedestrian is initially hidden behind another object).

Another significant number of pedestrian accidents happen at intersections, particularly when vehicles are turning.

Comparison with emergency-braking systems for vehicles traveling in the same direction

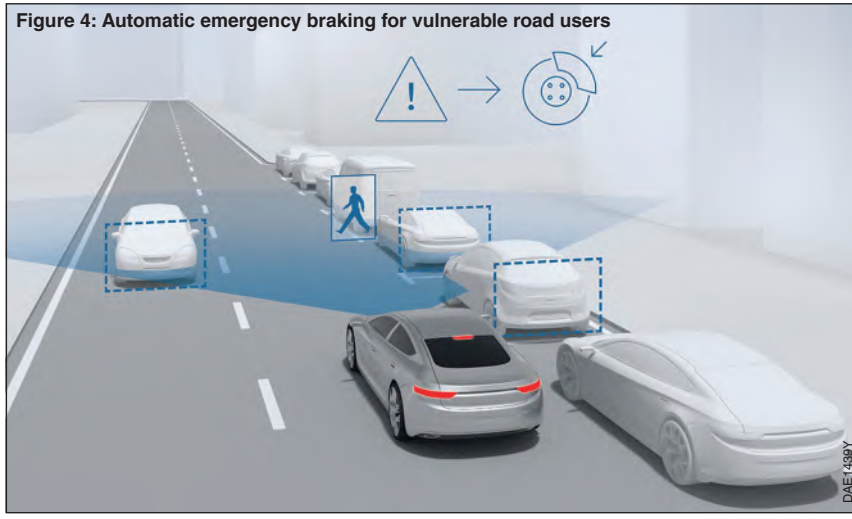
System responses

As already mentioned, the system responses for active pedestrian protection are comparable with emergency braking for vehicles traveling in the same direction. The system responds to an imminent collision with a pedestrian with a warning to the driver together with partial and full braking. As pedestrians frequently step out unexpectedly from a concealed position into the road, often only part of the response cascade, i.e., only the partial and full braking for example, is activated.

Environment detection

Pedestrians usually step out from the side into the road. Therefore, when compared with emergency-braking systems for vehicles traveling in the same direc-

Figure 4: Automatic emergency braking for vulnerable road users



tion, a wider area in front of the vehicle must be monitored by the sensor system. To make the correct decision on deployment, the sensor system must be able to measure precisely the lateral velocity of pedestrians in particular. Because pedestrians often step out onto the road from a concealed position, sensors must furthermore be able to detect those pedestrians very quickly (i.e., within a few sensor measurement cycles). This is the only way of ensuring a punctual system response (e.g., full braking).

Various sensor configurations for active pedestrian protection are in series production. In this case, both systems which are based on a single sensor and systems which fuse the information from a number of sensors are used. Sensors used are for example mono video cameras, stereo video cameras, or radar sensors.

Situation analysis

To assess the traffic situation, assumptions must be made about the future movement of the road users involved – i.e., here that of the pedestrian and of the driver's own vehicle. In this way, the system can calculate whether an imminent collision situation can still be averted by the driver or whether an activation of automatic emergency braking is necessary. This prediction is performed in all emergency-braking systems (e.g., emergency-braking systems for situations with vehicles driving in the same direction), but is particularly challenging in the case of active pedestrian protection. Pedestrians can very abruptly stop in their tracks, start to run, or change direction. These movement possibilities must be taken into account in the system in order to avoid unauthorized system interventions. For example, so as not to risk a rear-end collision with the following traffic, the system is not permitted to initiate emergency braking for a pedestrian who steps out onto the road but then stops in time.

Assessment standards

Since 2016 automatic emergency braking for pedestrians has been part of the Euro NCAP assessment (European New Car Assessment Program). Emergency-braking systems are assessed by reference to standardized collision scenarios which are recreated with dummy pedestrians on a test track. Three basic collision scenarios are tested, with the tested vehicle driving straight ahead at different speeds in all the cases:

– an unconcealed pedestrian crossing from the opposite side of the road (in right-hand traffic therefore from the left),
– a pedestrian crossing from the same side of the road (in right-hand traffic therefore from the right), and
– a pedestrian crossing from behind a parked vehicle.

In the last case the dummy pedestrian is the height of a child. The dummy pedestrians are able to move their legs and thus realistically mimic a real walking pedestrian. This is necessary since the movement of the legs is sometimes an important attribute in classifying the detected object as a pedestrian, without which automatic emergency braking would not be initiated. Radar sensors detect for example the leg movement of pedestrians via the micro-Doppler effect. This is characterized by a minimal oscillation of the pedestrian's measured speed signal.

Emergency braking for vulnerable road users traveling in the same direction

Rear-end collisions are among the worst – especially if pedestrians or cyclists are involved (Figure 5). The predictive pedestrian-protection system can prevent such collisions entirely or at the very least significantly mitigate their impact.

Up to 60 km/h the system can trigger automatic emergency braking in order to prevent frontal collisions with pedestrians or at the very last mitigate their impact.

Operating principle

In order to trigger automatic emergency braking before an accident occurs, the vehicle must know at all times if there are any objects in the danger zone in front of the vehicle. Radar sensors installed in the vehicle are on the one hand able to detect objects and other road users thanks to radar sensor technology with chirp-sequence modulation with a high level of reliability and accuracy, even in adverse weather conditions and poor visibility. On the other hand, cyclists and pedestrians can be detected promptly thanks to their wide aperture angle. In addition, the direction of travel and speed of the vehicle are constantly compared with the data of the pedestrian or cyclist. If the system detects a dangerous situation for pedestrians or cyclists, it can warn the driver and automatically initiate emergency braking if the driver fails to react. In this way, a collision with the endangered road user can be avoided or the vehicle's speed can at least be reduced as much as possible prior to impact if a collision is inevitable. This significantly minimizes the risk of serious injury.

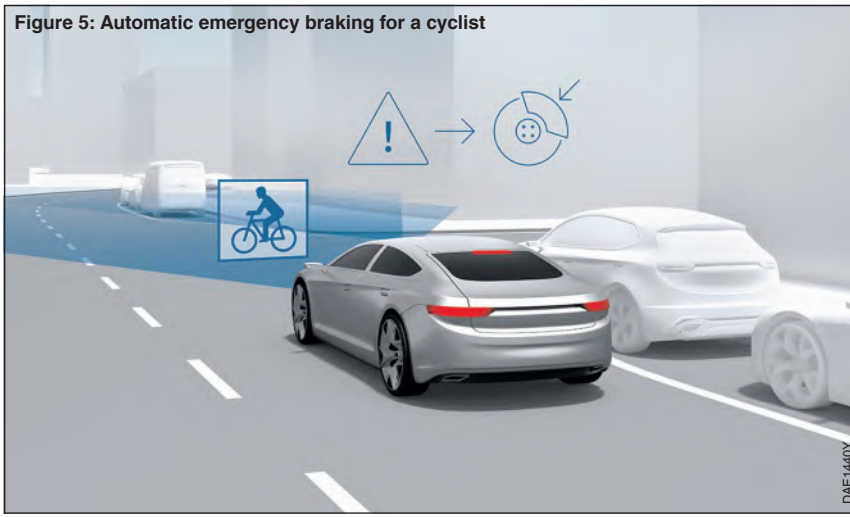
Reverse automatic emergency braking

Backing out of a parking space can be a challenge if the driver fails to adequately observe the surroundings behind the vehicle. A fleeting moment of inattention is enough to overlook a pedestrian passing behind a vehicle (Figure 6). Reverse automatic emergency braking for pedestrians reduces this risk as the system monitors the surroundings behind the vehicle and warns of an impending collision. If the driver does not react in time, the system automatically brakes the vehicle. This makes backing out of parking spaces safer and more comfortable.

Environment sensing

With the combined use of ultrasonic sensors and near-range cameras, the sensor data can be merged in the ECU and a 3D surround view created for a better and more robust understanding of the scene. Other road users, objects, and

Figure 5: Automatic emergency braking for a cyclist



parking-space markings can be detected and understood by means of data fusion. This makes parking functions and automatic emergency braking for vulnerable road users even more reliable and safer.

Thus, for example, while the vehicle is being backed out of a parking space, the system detects when a pedestrian is walking past behind the vehicle and relays this information to the brake control system, which then triggers automatic emergency braking.

Safety

The predictive pedestrian-protection system complies with the NCAP requirements regarding automatic emergency braking.

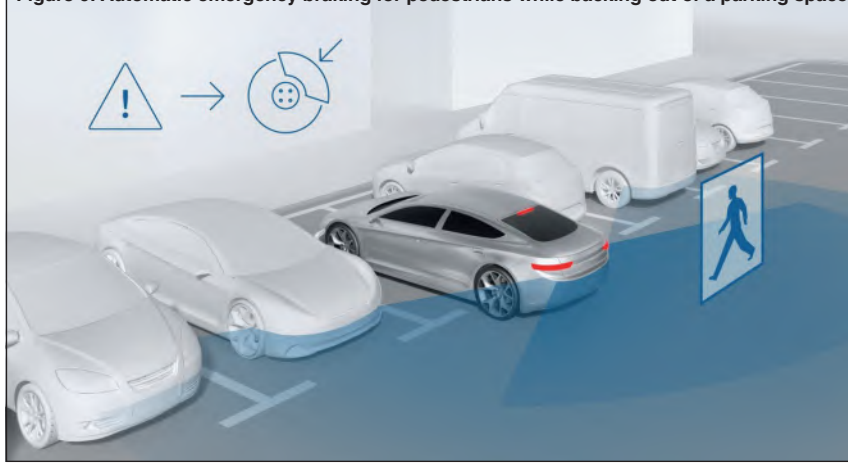
References

[1] A. Georgi, M. Zimmermann, T. Lich, L. Blank, N. Kickler, R. Marchthaler: New approach of accident benefit analysis for rear end collision avoidance and mitigation systems. In Proceedings of 21st International Technical Conference on the Enhanced Safety of Vehicles (2009).

[2] Commission Regulation (EU) 2015/562 of 8 April 2015 amending Regulation (EU) No. 347/2012 implementing Regulation (EC) No. 661/2009 of the European Parliament and of the Council with respect to type-approval requirements for certain categories of motor vehicles with regard to advance emergency braking systems.

[3] UN-ECE R131: Regulation No. 131 of the Economic Commission for Europe of the United Nations (UNECE) – Uniform provisions concerning the approval of motor vehicles with regard to the Advanced Emergency Braking Systems (AEBS).

Figure 6: Automatic emergency braking for pedestrians while backing out of a parking space



Intersection assistant

Motivation

Cause of accidents
Around 26 % of all accidents involving injury in Germany occur when crossing intersections and turning [1]. Driver-assistance systems which are designed to prevent this common type of accident are called intersection assistants. This covers safety functions with the aim of avoiding accidents or mitigating their severity. To achieve this, the driver is alerted at an early stage to an imminent accident or an automatic braking of the vehicle is performed.

Environment sensing
Other road users are detected with radar sensors, video cameras or lidar sensors and their distance, speed and direction of movement are determined. The information is processed by the microcomputer of an ECU and evaluated in the form of an object list in software.

Situation assessment
An imminent accident is predicted whereby the future position of the vehicles is predicted from the measured positions and speeds of other road users and the motion status of the driver's own vehicle. In the next step the system checks whether the predicted vehicles coincide, which would correspond to an imminent collision. It is however necessary to take into account in this respect that the measured positions and speeds are invalidated by measuring errors. It is furthermore not known exactly how the vehicles will move in future, i.e., whether they will for example brake or steer. Therefore, when it comes to predicting the situation development, driving maneuvers which would avoid an accident must also be taken into account. The less these avoidance maneuvers are possible, the more critical the situation. The situation where an accident cannot be avoided or can only be avoided with difficulty by means of normal handling is called an acute collision risk.

It is particularly difficult to distinguish between critical and non-critical situations in intersection scenarios as the vehicles have a large freedom of movement in several direction and the turning behavior of the vehicles involved would have to be predicted several seconds in advance. This is almost impossible, especially when the vehicles involved are being driven in a sporty style. For this reason, an imminent accident can only be unequivocally detected just before the collision.

Unauthorized brake interventions must be avoided without fail as firstly they would not be accepted by the driver and secondly they constitute a potential danger to the following traffic. In certain situations, therefore, automatic emergency braking cannot be activated or can only be activated with a delay. Accident prevention is then no longer possible in some circumstances. If, for example, the driver's vehicle is traveling at high speed, an accident-avoiding braking operation would as a rule have to be initiated very early, but this is not possible due to the situational uncertainties and the risk of incorrect deployments.

But even if the system cannot avoid an accident in an individual case, it can mitigate the consequences of the accident through an automatic braking intervention since the impact speed is reduced and the point of impact favorably shifted under certain circumstances.

A distinction is made between the intersection-assistance systems described in the following.

Left-turn assistant

Causes of accidents
The left-turn assistant is designed to prevent accidents which happen when turning left in the face of oncoming traffic (Figure 1). The accident is usually caused by the driver incorrectly judging the speed of the oncoming vehicle – the accident perpetrator turns although there is no longer enough time to complete the maneuver.

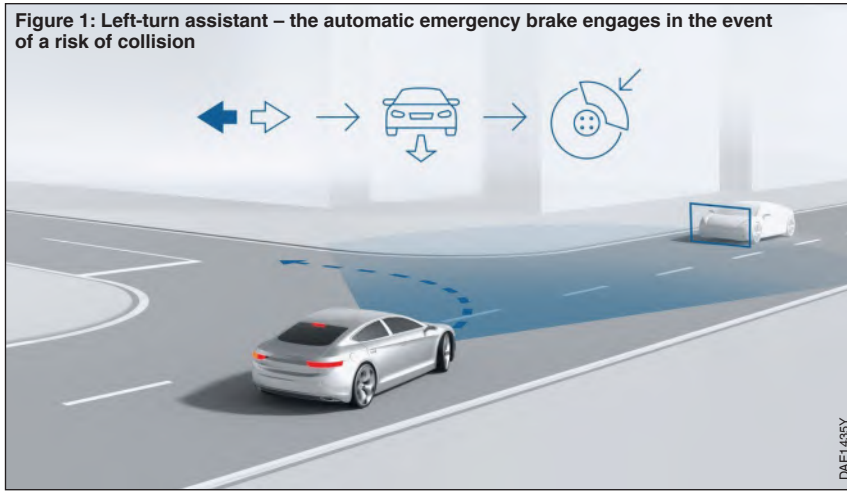
Environment sensing
The multifunction camera used in addition to the front radar sensor helps to increase the availability of the AEB functions (automatic emergency braking). Objects can be detected, classified, and tracked more reliably thanks to the fusion of the sensor data from the radar and the camera. This creates an extremely detailed and reliable "image", which is the basis of a robust interpretation of the vehicle's surroundings.

System response
When turning, the driver must not only keep an eye on crossing traffic, but also on oncoming traffic. Disregarding or misjudging the speed of an oncoming vehicle can quickly develop into a dangerous situation (Figure 1). Here automatic emergency braking engages for turning scenarios. An imminent left-turn maneuver by the driver can be detected by an evaluation of turn signal, accelerator-pedal position, and steering angle.

When the vehicle is stationary in preparation for making a turn and the driver tries to make the turn even though there is a real danger of collision, the driver is warned in a first stage. If there is an acute risk of collision, automatic braking or start prevention is finally initiated. The driver is prevented from moving off until the danger has passed. If the vehicle has already been set in motion despite the risk of collision, the system recognizes the dangerous situation and warns the driver. The driver is then responsible for reacting appropriately.

To avoid the situation where the vehicle in the event of an incorrect response cannot be moved in a potentially dangerous situation, the driver can override the start prevention by means of a kickdown.

Figure 1: Left-turn assistant – the automatic emergency brake engages in the event of a risk of collision



Cross-traffic assistant

Causes of accidents
The cross-traffic assistant is designed to prevent accidents which happen when crossing intersections and turning in the face of vehicles that are themselves crossing (Figure 2). Because passengers involved in side collisions are much less adequately protected than in head-on collisions due to the lack of crumple zones, the risk of injury is particularly high.

Obscurations/obstructions are frequent causes of accidents as well as distracted and inattentive drivers.

Environment sensing
As with the left-turn assistant, crossing vehicles are detected by environment sensors. Compared to the left-turn assistant, however, the field of vision to be covered is much wider (≥180°), meaning that generally several sensors must be used. Radar sensors mounted at the front corners of the vehicle enhance the vehicle's horizontal field of vision and enable the system to detect crossing vehicles even earlier and faster, even those traveling at higher speeds. These radar sensors are ideally designed for detecting and differentiating between relevant objects, particularly in hectic and confusing urban traffic.

System response
If while the vehicle is entering a blind intersection the system detects a crossing vehicle approaching, the driver is alerted in an initial stage to the presence of a potentially dangerous traffic situation (e.g., visually by way of a neutral indication in the head-up display). If the driver attempts to enter the intersection even though another vehicle will shortly be crossing the lane and a safe crossing of the intersection is no longer possible, the system prevents the vehicle from moving, for example by cutting the fuel supply or engaging the brakes. As with the left-turn assistant, start prevention can be overridden by means of a kickdown. At faster speeds a warning (conspicuous visual or acoustic indications) is issued first in the event of an imminent collision with cross-traffic. If the driver fails to react, the system initiates an automatic brake intervention in the event of an acute collision risk.

Outlook
Because obscurations/obstructions caused for example by parked vehicles are frequently encountered in intersection situations, presumably car-to-car communication (C2C) will also play an important role in the future. Here the vehicles exchange information by radio signal on their position, speed, and steering angle, and can thus detect potential conflicts at an advanced stage and take countermeasures. The position of the vehicle is determined beforehand by satellite positioning. However, extremely accurate locating of the vehicles is required to enable the information from a safety-critical function like the cross-traffic assistant to be used. This is not possible however with the precision of GPS that is currently used by commercially available navigation devices. This is caused by for example atmospheric interference. In the case of Differential GPS (DGPS) a second GPS receiver, whose exact position is known, is used. This generates a correction signal which can be used by other GPS units in the immediate surroundings to correct the interference. Centimeter-perfect locating is thus theoretically possible. However, this system is currently still very expensive and therefore cannot be used for series introduction in vehicles. Furthermore, the GPS receiver must be in direct radio contact with a larger number of satellites, which is often not the case in urban canyons.

Traffic-light and stop-sign assistants

Causes of accidents
Unlike the left-turn and cross-traffic assistants, the traffic-light and stop-sign assistants are designed not to prevent an imminent collision with other road users, but to prevent the driver from accidentally running a red light or a stop sign. This is intended to prevent accidents which happen as a result of these traffic violations.

Environment sensing
A video camera equipped with image-recognition algorithms scans the vehicle environment for traffic lights and road signs (see Road-sign recognition), whereupon their status and their meaning are interpreted. To reduce the risk of erroneous detections, digital maps can be used in which the positions of road signs are marked.

An alternative is offered by car-to-infrastructure communication (C2I). Here, the traffic light for example communicates its status by radio signal to vehicles equipped with special receivers in the surrounding area. Furthermore, the associated stop line must be detected to enable the system response to be sensibly configured in terms of time.

System response
If the driver is on the point of not stopping at a red light or a stop sign, a warning is issued first. If the driver then fails to react appropriately, the vehicle can also, depending on the system design, be automatically braked.

A sensible response to the yellow phase is only possible if there is still sufficient distance for a comfortable stopping operation or if the remaining duration of the yellow phase is known for example via C2I.

Another type of traffic-light assistance pursues the objective of optimally utilizing the green phases and avoiding unnecessary braking and acceleration phases. Aside from the fluid and comfortable driving style, this is also intended to lower fuel consumption. In this system the traffic lights ahead transmit their traffic-light phase and the time until the phase change to approaching vehicles. The speed at which the next traffic light is reached on green is displayed in the instrument cluster.

References
[1] GIDAS database 2014.

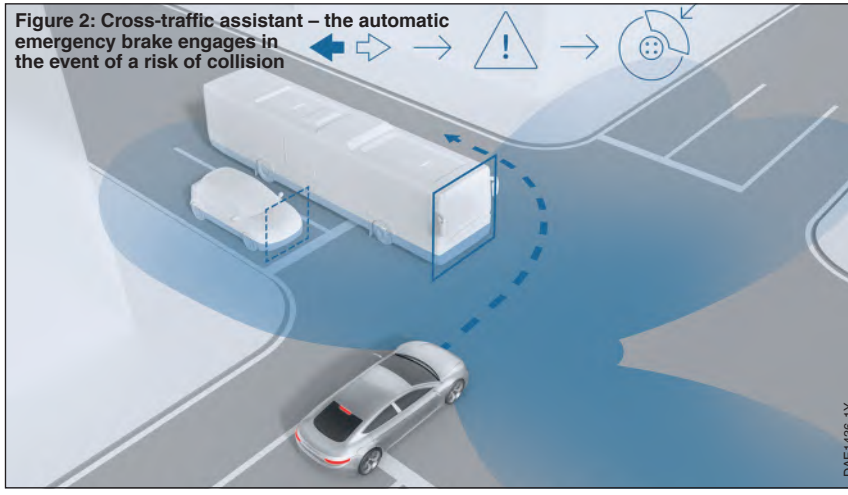


Figure 2: Cross-traffic assistant – the automatic emergency brake engages in the event of a risk of collision

Intelligent headlamp control

Motivation

Statistics show that people are exposed to a significant risk of being involved in a serious accident when driving at night. A typical driver when driving at night for the most part does not switch on the high beam and drives on low beam even though activating the high beam is perfectly possible. Some drivers do not engage the high beam at all and continuously drive on low beam. A system with a high-beam assistant activates the high beam much more frequently than the typical driver who activates manually and thereby significantly improves the illumination of the road compared to the low beam. The figures from the statistics are very much dependent on the traffic density – the lower the density, the greater the use of the high beam. It has also been established that fatigue and lack of concentration causes drivers not to manually turn off the high beam reliably and thus to dazzle other road users.

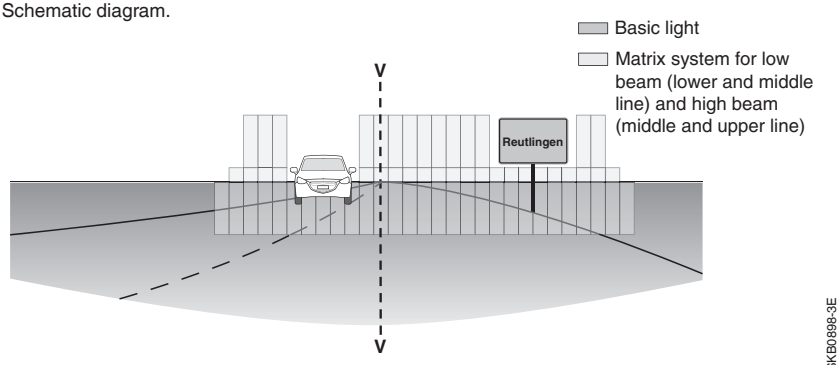
An automatic system offers a dramatic improvement for all road users. The driver of a vehicle with a high-beam assistant experiences a noticeable reduction of the burden and, thanks to automatic activation, a more frequent improvement in illumination. Automatic deactivation of the high beam ensures that drivers of oncoming vehicles are not dazzled.

System designs

The high-beam assistant automates the process of switching between low beam and high beam. It detects vehicles driving in front and oncoming vehicles and automatically switches the high beam on or off to suit the traffic situation. This significantly increases safety and comfort when driving at night, mainly on highways and expressways.

Basic version
The system consists of a light-sensitive sensor or a camera, combined with an ECU which controls the headlamps. The camera is either integrated in the rearview mirror or installed in a standalone sensor assembly on the windshield. The camera detects and distinguishes between vehicles and stationary street lights. The system detects oncoming vehicles up to a distance of 600 m and tail lamps of vehicles driving in front up to a distance of 400 m. The positions of detected objects are transmitted to the headlamp ECU, with the data being transmitted via the CAN bus used in the vehicle. The headlamp ECU calculates from the position data whether an object is in the headlamp's beam range and dazzling can occur. The high beam is automatically deactivated if the sensor system detects an object that can be dazzled. If the sensor system does not detect an object, the high beam is automatically activated. The high beam is deactivated even in the presence of street lights for example in urban areas which the system detects and can distinguish from other objects.

Figure 1: Matrix headlamps
Schematic diagram.



Further expansion stages
The range of the low-beam or high-beam lights can be automatically adjusted using video data.

The next stage of intelligent headlamp control facilitates adaptive high-beam control. This not only controls the range of the light, but also adapts the width of the beam to the traffic conditions. This can, for example, illuminate curves in advance or in urban traffic deliver a wider light cone which can effectively illuminate the edges of the road, ultimately helping the driver to spot any potentially vulnerable pedestrians.

With a further expansion stage of headlamp control the driver can drive with the high beam permanently on. Oncoming traffic or traffic ahead and its position are detected on the basis of camera-image analysis. The system is based on horizontally and vertically swiveling headlamps or full LED headlamps in which the entire light distribution is controlled in segments so that on the one hand road users who could potentially be dazzled are not exposed to the light cone (matrix headlamps, Figure 1), while on the other hand the remainder of the area can be optimally illuminated by the high beam (“Glare-free high beam”). The light distribution from the high beam remains virtually unchanged, while the driver's visual range is increased considerably. Most of the high beam can thus remain switched on and contribute to the driver’s safety and comfort without dazzling other road users.

Advanced rider-assistance systems for two-wheelers

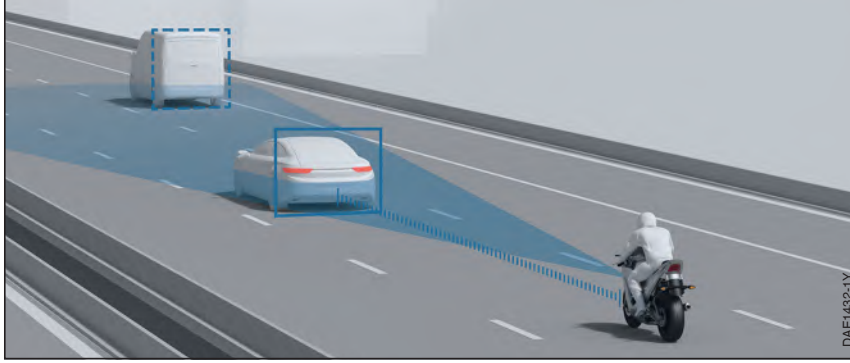
Assistance systems such as the antilock braking system (ABS) and motorcycle stability control (MSC) have already made riding on two wheels considerably safer. Advanced rider-assistance systems (ARAS) can make a further contribution to increasing safety. The electronic assistants are always vigilant and, in emergencies, they respond more quickly than humans can. The technology underpinnings these systems is a combination of radar sensor, braking system, engine management, and HMI (Human Machine Interface). Giving motorcycles radar a "sensory organ" makes these new assistance and safety functions for two-wheelers possible while providing a accurate picture of the vehicle's surroundings. As a result, these assistance functions not only increase safety, they also enhance enjoyment and convenience by making life easier for riders.

– With ARAS the rider can concentrate better on the current traffic situation,
– the system supports the rider in critical situations and helps to avoid accidents,
– more comfortable and relaxed riding,
– the electronic assistants are always vigilant and, in emergencies, they respond more quickly than humans can.

Adaptive cruise control

Riding in heaving traffic and maintaining the correct distance to the vehicle in front takes a great deal of concentration and is exhausting over longer periods. Adaptive cruise control (ACC) adjusts the vehicle speed to the flow of traffic and maintains the necessary safe following distance to the vehicle in front (Figure 1, see also Adaptive cruise control). ACC can effectively prevent rear-end collisions caused by insufficient distance to the vehicle in front. And not only does ACC offer riders more comfort and convenience, for example when riding in convoy, it also allows them to concentrate more on the current traffic situation.

Figure 1: Adaptive cruise control
The motorcycle adapts the speed to the vehicle in front.



Forward collision warning

In road traffic, even the briefest lapse in concentration can have serious consequences. The forward collison warning for motorcycles reduces the risk of a rear-end collision or at the very least mitigates it impact. The system is active as soon as the vehicle starts and it supports the rider in all relevant speed ranges. If the system detects that a vehicle in front is dangerously close and the rider fails to react to the situation, it warns the rider by way of an acoustic or visual signal (Figure 2).

Forward collision warning alerts the rider promptly to the danger of a rear-end collision so that the rider is able to react in time.

Blind-spot detection

Blind-spot detection helps motorcycle riders change lanes safely. A radar sensor acts as the system's electronic eye, detecting objects in hard-to-see areas. Whenever there is a vehicle in the rider's blind spot, the technology warns the rider with a visual signal, for example in the rearview mirror (Figure 3).

Figure 2: Forward collision warning
A number of vehicles are detected by the environment-sensor system, danger comes only from the vehicle traveling in front in the same lane.

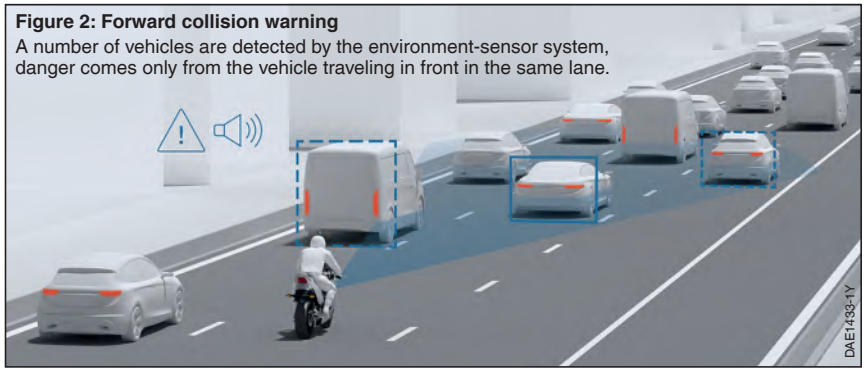
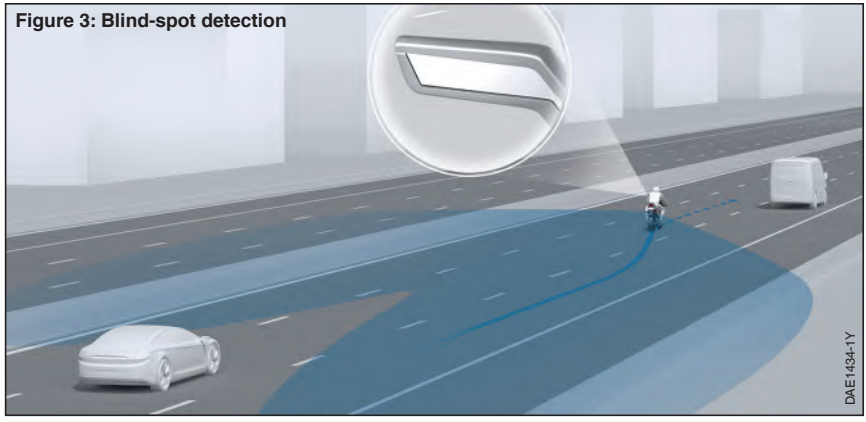


Figure 3: Blind-spot detection



Interior-monitoring systems

Interior monitoring enhances safety, comfort, and convenience for all vehicle occupants. The system for monitoring the vehicle interior from Bosch is able to detect signs of distraction or drowsiness and even whether a child has been left behind in the vehicle, and alerts the driver to critical situations. Safety systems, such as for example the seat-belt alert function, are further enhanced by information acquired from the vehicle interior. Besides the various safety-relevant applications, the system also provides a range of innovative interaction options, like gesture control for distraction-free operation of the infotainment system. Vehicle interior-monitoring systems will in future be absolutely essential in automated vehicles for ascertaining whether the driver is able is able to take control of the vehicle in a particular situation.

Camera-based interior monitoring

The modular system provides vehicle manufacturers with a high degree of flexibility when adapting the system to different requirements and vehicle types.

It means the tailored solutions can be implemented quickly and cost-effectively, and it allows other systems, such as automated driving, passive safety, and infotainment, to be integrated seamlessly.

Driver monitoring

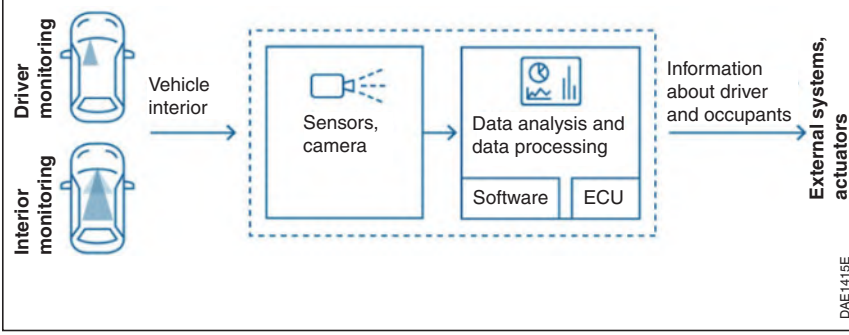
Driver-drowsiness detection

The driver-monitoring camera (Figure 1) monitors the alertness and state of the driver. It can detect distraction, drowsiness, and microsleep and alerts the driver promptly.

Driver identification

Facial recognition makes it possible for the system to identify the driver with absolute certainty. Using the stored driver profile, the system can then automatically adjust personal comfort and convenience settings such as optimal seat and mirror positions, favorite radio station, or preferred interior temperature.

Figure 1: Camera-based interior monitoring



Interior and occupant monitoring
With the interior-monitoring system, it is not just the driver who is the focus of attention. The camera is positioned in such a way that all seats are in its field of vision. The system can detect the presence of any other occupants and can then provide more specific reminders to fasten seat belts.

The system is also able to identify whether the front passenger seat is occupied and can thus deactivate the airbag if, for instance, a child safety seat is present.

Rising temperatures in the vehicle can quickly become dangerous for a child left behind on the back seat. The system warns the driver in this case too in order to prevent a critical situation like this occurring in the first place.

Safety and convenience
The interior-monitoring system makes it possible to implement innovative interaction interfaces, such as gesture control. This enables the driver, for instance, to skip to the next song with a simple hand movement without looking away from the road. This reduces distraction and at the same time enhances the “user experience”.

Driver monitoring for automated driving
In future automated vehicles it will be necessary to know whether the driver is able in critical situations to assume control of the vehicle again.

In automated driving situations the driver can adopt a more relaxed posture because they no longer have to control the vehicle permanently. In dangerous situations the system can determine the seat position and move the occupant into a safe position so that they are afforded optimal protection in the event of a collision.

Future of automated driving

Automation levels

Driving a car can be fun, but can also be tiring or even dangerous. Driver-assistance systems support drivers in difficult driving situations or relieve them of tiresome and monotonous driving tasks.

A trend that has been observed in recent years is the increasing degree of automation of the systems on the market. This trend will continue in the future [1], [2].

Levels according to SAE
The definitions of the SAE (SAE J3016 [3]) have gained acceptance as the standard worldwide when it comes to classifying the driving functions of continuous automation. There are five different levels.

Level 1
An assistance system assumes lateral or longitudinal control of the vehicle in a specific application. The driver is supported, but remains responsible.
Examples: Lane-keeping systems, adaptive cruise control (ACC).

Level 2
The system assumes longitudinal and lateral control in a specific application. The driver is supported, but remains responsible. The driver monitors the system constantly and takes on the delegated tasks where necessary.
Example: Highway assistant.

Level 3
The system takes on the entire driving task in a specific application. The driver no longer has to monitor the system constantly, but must be able to take over the driving task again at any time within a short period of time following a takeover request.
Example: Traffic-jam pilot section.

Level 4
The system takes on the entire driving task in a specific application. In the event of a fault, responsibility remains with the system, which brings about a safe state; it does not send a takeover request to the driver.
Examples: Driverless city taxi, automated valet parking.

Level 5
The system takes over the entire driving task in all applications, for instance in cities, on ordinary roads or expressways/highways. The user simply inputs a destination and the vehicle navigates to this destination all by itself. In the event of a fault, responsibility remains with the system, which brings about a safe state.
Example: Vehicle can drive autonomously at any time and under all conditions.

Autonomous driving
The logical extension of this development is therefore a system that can permanently perform all the driving tasks without the driver having to monitor the system. The driver becomes a passenger and can use the journey time for other activities, such as for example reading, working or even sleeping.

Further applications would include the vehicle independently looking for a parking space (see Automated valet parking), driving to the garage or taking the children to school without a driver actually having to be in the vehicle. Last but not least, it is hoped to reduce the number of accidents in that a computer-controlled system can essentially react faster and more reliably than a person and unlike the driver does not get tired.

Obstacles to automated driving

The idea of driverless driving has been pursued for several decades in numerous research projects. As early as 1986 Professor Dickmanns (University of the Federal Armed Forces, Munich) introduced a “robot vehicle” which was able to drive under automated conditions on expressways/highways at speeds of up to 96 km/h. In 1995 Professor Dickmanns’ team drove another vehicle from Munich to Copenhagen and back (covering a distance of 1,758 km) with automatic longitudinal and lateral guidance. The system was available for 95 % of the journey time.

In 2004 the US Department of Defense organized the DARPA Grand Challenge (DARPA, Defense Advanced Research Projects Agency), the first competition over a longer distance for autonomous vehicles.

The DARPA Urban Challenge organized in 2007 involved the autonomously driving vehicles having to perform a range of tasks in quasi-urban traffic in a former air-force base.

As the various research projects have shown, today it is perfectly possible to build a Level 4 or Level 5 prototype vehicle. But before such vehicles can be put into series production, some technical and non-technical obstacles have to be overcome.

Since 2011 Robert Bosch GmbH has been joining various vehicle manufacturers and other suppliers in working on the technologies that are required for automated driving. Since 2013 Bosch has also been running test vehicles in public traffic.

Technical obstacles

The technical aspects involved are, from a functional standpoint, essentially environment detection, understanding of the situation, and the behavior planning derived from them.

Environment detection

With regard to environment detection, all aspects of the vehicle’s surroundings as are a matter of course to a human driver must be sensed and detected. These aspects include detecting other road users and detecting the road, obstacles that cannot be driven over, and the traffic infrastructure (e.g., road markings, road signs and traffic lights).

Understanding of the situation

In the final analysis, the gestures of other road users and traffic police officers must also be interpreted. Both the reliable measurement at all times of all these aspects and the processing of the data supplied by these environment sensors are currently still stretched to their limits.

Behavior planning

A further functional challenge is to derive appropriate actions (behavior planning) in real time. Traffic situations are often highly complex, particularly in urban environments, due to the fact that a multitude of road users, traffic regulations and special situations has to be taken into account. A number of pilot projects are already addressing the urban sphere, e.g., in the low speed range. These systems are however not yet available in production vehicles.

Outlook

With the inexorable advances in development (e.g., with increasing computing capacity, powerful learning algorithms) it appears that these functional obstacles can be surmounted in the foreseeable future.

As for a further technical obstacle, the validation of functional reliability, there is as yet no general solution approved by legislators for series production: A fully automated vehicle naturally may not jeopardize its occupants or other road users in any situation through failures/faults. The failure rate will therefore have to satisfy extremely stringent requirements for a

series-production launch. To ensure this, many millions of kilometers would have to be completed in test drives, which is currently not economically feasible and necessitates the use of simulations.

Legal obstacles
Finally, there are also legal (non-technical) obstacles that need to be removed. It is unclear for instance who is liable for the damage/injury caused by a vehicle at automation level 3 or higher when the driver no longer has responsibility for vehicle handling.

Aside from the question of liability, Level 3 and higher automated driving functions in public traffic are currently not legal in many countries. Often there are no binding approval criteria for vehicles with systems at higher automation levels. Furthermore, the regulatory legislation needs to be adapted because driving strategy and behavior in traffic are determined no longer by the driver, but by the algorithms of a technical system.

The responsibilities for the legislation are regulated differently worldwide, resulting in a fragmentation of the requirements. The UNECE (United Nations Economic Commission for Europe) has a central role in the goal of harmonizing the requirements.

Stages on the road to autonomous driving

Autonomous driving (Level 5) will not be introduced suddenly, but instead will come in stages. In other words, initially systems at a lower automation level (Level 2), which are confined to specific use-cases, will be brought into series production. These can be for example:
– expressway/highway situations, due to the more simply structured and less complex environment,
– traffic-jam, parking and maneuvering situations, since they typically occur in the low-speed range, the sensors involved only need to have a short range, and the vehicle can be brought to a stop very quickly and with only a minimal stopping distance.

Combinations are naturally also conceivable, such as the traffic-jam pilot described in the next section.

Development of assisting driving functions
Assisting driving functions such as adaptive cruise control (ACC), lane-keeping support, and parking assistant have been established on the market for some years now. And also Level 2 functions such as the traffic-jam assistant are now available on the market.

Traffic-jam assistant
Here the system performs lateral and longitudinal guidance completely in traffic-jam situations. At low speeds the driver depending on the functionality does not have to keep his hands on the steering wheel, but does have to monitor the system and intervene if necessary. The traffic-jam assistant as a rule functions in a speed range up to 60 km/h.

Highway assistant
The highway assistant assumes longitudinal and lateral control at typical expressway/highway speeds (e.g., combined system of ACC and lane-keeping support). But ultimate responsibility remains with the driver. He must monitor the system so that he can intervene at any time. Depending on the functionality, the highway

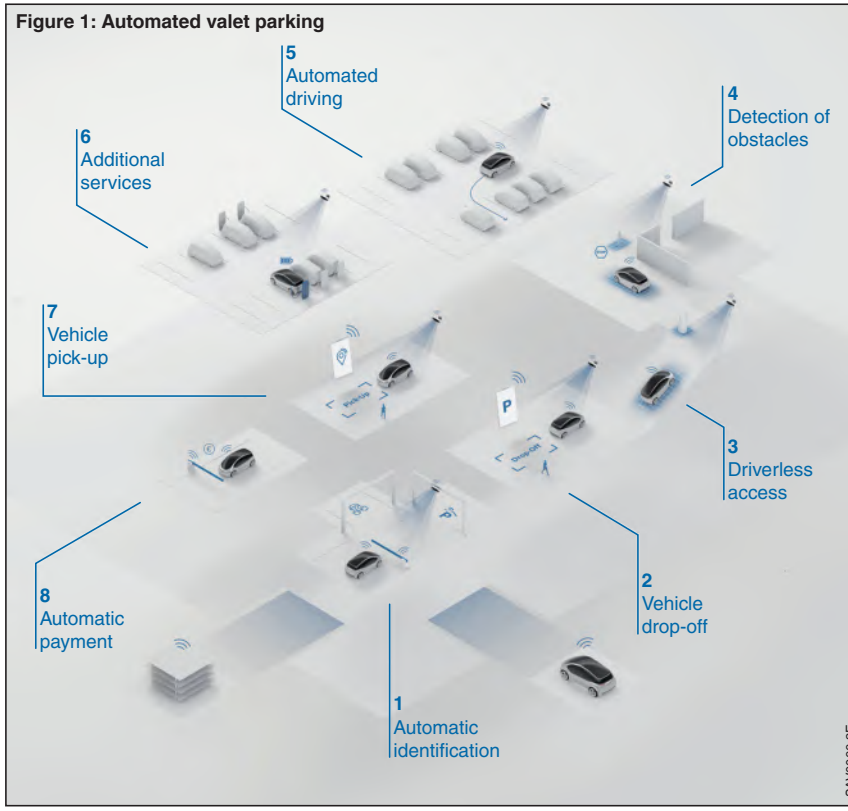
assistant can also contain the traffic-jam assistant function.

Traffic-jam pilot and highway pilot
The next level of automation (Level 3) includes for example the traffic-jam pilot, which assumes complete responsibility for vehicle handling/guidance in traffic jams. To realize the follow-on highway pilot, it is necessary to provide the emergency stop subfunction, which can bring the vehicle to a safe stop on the hard shoulder (emergency lane) if the driver is unable to re-assume responsibility for driving. Unlike the highway assistant, the highway pilot assumes responsibility for vehicle handling/guidance. The driver can perform secondary activities, but must where necessary be able to take over the driving task from the system again within a few seconds.

Automated valet parking
On top of the expressway/highway functions, the parking and maneuvering functions are also subject to ever-increasing automation. Systems which drive the vehicle independently under human supervision into a parking space are already available (see Parking and maneuvering systems).

Automated valet parking goes one step further. A smart parking-garage infrastructure works hand in hand with the technology in the vehicle to handle the entire parking operation in the parking garage under fully automated conditions (Figure 1). The necessary parking-garage infrastructure includes cameras which detect vacant parking spaces in the parking garage, monitor the driving corridor and its surroundings, and detect unexpected obstacles or persons in the lane so that

Figure 1: Automated valet parking



the vehicle can react immediately. The technology in the vehicle converts the infrastructure's commands into driving maneuvers. The data acquired by the cameras are processed by the parking-garage IT technology to compute the routes and satisfy all the safety requirements. A technical unit which communicates with the vehicle and a cloud connection for interacting with the back end are also required.

Once the driver has parked their vehicle in a drop-off area in front of the parking garage, the vehicle is safely and efficiently navigated by the system into a vacant parking space. The parking operation comprises the following individual stages (see Figure 1):

1. Automatic identification
Because an “automated valet parking” space has already been reserved through an app on the smartphone or by the in-car infotainment system, the smart infrastructure already identifies the vehicle at the entrance.

2. Simple vehicle drop-off
The driver comfortably parks their vehicle in the drop-off area, gets out, and activates the valet service by smartphone command. From this moment on, the “automated valet parking” infrastructure is in constant communication with the vehicle and navigates the vehicle to a vacant parking space.

3. Driverless access
The cameras installed in the parking garage detect vacant parking spaces and monitor the driving corridor and its surroundings.

4. Detection of obstacles
If unexpected obstacles such as persons or objects are detected in the surrounding area by the smart “automated valet parking” infrastructure, the vehicle is immediately brought to a stop.

5. Automated driving
The vehicle maneuvers itself into the tightest parking space thanks to the “automated valet parking” infrastructure.

6. Additional services
Automated valet parking extends the range of possible services, such as for example driverless battery charging while the vehicle is parked in the parking garage.

7. Simple vehicle pick-up
As soon as the driver is ready to pick up, they activate their vehicle by smartphone and the smart parking-garage infrastructure directs the vehicle automatically to the pick-up area.

8. Automatic payment
Payment is taken automatically at the exit. Depending on the business model, all the partners involves share the parking charge.

Bosch and Daimler have in this connection reached a milestone on the road to automated driving. In summer 2019 the two companies received the approval of the relevant authorities for the automated parking system in the parking garage of the Mercedes-Benz Museum in Stuttgart/Germany. It is therefore the world's first fully automated and driverless parking function in compliance with SAE Level 4 officially approved for everyday use.

Long-term goal – autonomous driving
The logical extension of development is an automation level 5 vehicle that can take the passengers from door to door like a taxi. Here the vehicle must be able to drive autonomously at any time and under all conditions.

Such fully automated systems are also already in development. They are, on account of the absence of driver supervision and thus the absence of a human fallback level, based on a completely different system architecture from classic driver-assistance systems.

Thus Level 4 and 5 automated and autonomous systems must have a redundant design. As well as a failsafe power supply for safety-relevant system components and redundant actuators (e.g., steering intervention by power steering or asymmetrical intervention of driving-dynamics control), redundant sensor concepts must be used. The goal is therefore to compensate for technological weaknesses in the

various sensors. For example, a video sensor as an individual sensor can be dazzled by bright sunlight. A redundant sensor concept is based accordingly on the combined use of sensors that use independent measuring principles (e.g., radar, video, and lidar), but also overlap with regard to their sensing range.

In addition, a new central system component is introduced in fully automated systems with high-precision digital maps. This is required to provide information for representing the surrounding environment which cannot be measured with environment sensors (e.g., course of a road after the bend/curve).

Another new component is the behavior planner, which assumes the job of tactical maneuver planning which was previously performed by the human driver. Examples are the recognition of the need for a lane change, a slipping-in maneuver or an overtaking maneuver, to name just a few.

Exact vehicle locating is necessary to enable the data supplied by the environment sensors to be correctly associated with the digital map. The purpose of locating is, on the basis of the GPS position (Global Positioning System), the vehicle-motion data (e.g., the driving speed) and striking environmental features (e.g., trees, houses, city-limit signs), to pinpoint the exact position of the vehicle on the digital map. This is the only way of ensuring correct data association between environment sensors and map data.

As well as the pure measurement of stationary and moving objects in road traffic, a fully automated system must be able to understand and utilize all the paths of communication between road users (e.g., gestures and eye contact with a pedestrian on a zebra crossing, hand signals by cyclists, traffic signs carried by police officers or road workers). While technical solutions for communication between vehicles and infrastructure already exist with Vehicle2X, research is currently still being conducted into robust solutions to gesture recognition in road traffic.

The development of these systems is constantly progressing. In a few years these systems will leave their prototype status behind and will be adopted in our everyday lives.

References

[1] H. Winner, S. Hakuli, F. Lotz, C. Singer: Handbuch Fahrerassistenzsysteme – Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort. 3rd Ed., Verlag Springer Vieweg, 2015.

[2] M. Buehler, K. Iagnemma, S. Singh: The DARPA Urban Challenge: Autonomous Vehicles in City Traffic (Springer Tracts in Advanced Robotics). Springer-Verlag, 2010.

[3] SAE J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles.

Technical terms

Symbols

0.2 % yield strength,
material parameters 260
12 V starter battery,
automotive electrical systems 1478
12 V vehicle electrical system,
automotive electrical systems 1450
12 V vehicle electrical systems,
vehicle electrical systems
for commercial vehicles 1467
14 V vehicle electrical system,
hybrid and electric vehicles 1562
15° tapered bead seat rim,
wheels 1158
1-plunger radial-piston pump,
common rail 939
2/2-way solenoid valves,
wheel-slip control systems 1265
24 V vehicle electrical systems,
vehicle electrical systems
for commercial vehicles 1467
2-channel system,
antilock braking system 1268
2-day diurnal test,
evaporative-emissions test 1088
2-plunger in-line piston pump,
common rail 942
2-plunger radial-piston pump,
common rail 939
2-stroke cycle 666
360° all-around visibility,
parking and maneuvering
systems 1887
3-channel system,
antilock braking system 1268
3D point cloud,
lidar-sensor system 1825
3D surround view, information
and warning systems 1897
3-day diurnal test,
evaporative-emissions test 1088
3-part circular-key sectional profile,
frictional joints 436
3-plunger radial-piston pump,
common rail 937
3-sensor system,
antilock braking system 1268
48 V-12 V voltage transformers,
DC-DC converters 1596
48-hour diurnal test,
evaporative-emissions test 1088
4b/5b encoding of data,
Serial Wire Ring 1711
4-channel system,
antilock braking system 1267
4-stroke cycle 664
5° tapered bead seat rim, wheels 1158
α-decay, elements 186
β-decay, elements 186
γ-decay, elements 186
γ-radiation, elements 186
λ control loop,
three-way catalytic converter 879
λ control range,
three-way catalytic converter 875
λ control, Electronic Diesel Control 913
λ control,
three-way catalytic converter 879
λ mean value,
three-way catalytic converter 875
λ sensors 1766
λ/4 resonators,
exhaust-gas system 797
μ split, wheel-slip control systems 1262
σ level, technical statistics 230

A

A/C unit 1392
A/F (Air/fuel) mixture,
spark-ignition engines 838
A-arm links, wheel suspensions 1151
A-bracket, wheel suspensions 1148
ABS control loop,
antilock braking system 1261
ABS control methods,
antilock braking system 1263
ABS control,
wheel-slip control systems 1261
ABS controller, ESC 1278
ABS enhanced,
Motorcycle Stability Control 1300
ABS system variants,
antilock braking system 1267
ABS systems for commercial vehicles,
wheel-slip control systems 1269
ABS systems for passenger cars,
wheel-slip control systems 1265
Absolute frequency,
technical statistics 226
Absolute value function,
mathematics 206
Absolute-pressure sensors 1756
Absorbent glass mat,
starter batteries 1484
Absorber, vibration absorber 64

Absorption material, muffler 795
Absorption method,
diesel smoke-emission test 1085
Absorption mufflers,
exhaust-gas system 795
Absorption-type mufflers,
air-intake system 758
Ac1 temperature, heat treatment 334
ACC Stop & Go,
Adaptive Cruise Control 1894
Acceleration sensors 1750
Acceleration, mechanics 46
Acceleration, motor-vehicle
dynamics 472
Accelerator-pedal sensor 1725, 1731
Acceptance checking, CAN 1694
Access authorization,
locking systems 1431
Access methods, vehicle
communication networks 1686
Accessory Control, Motronic 804
Accessory drive, belt drives 723
Accident database,
driver-assistance systems 1805
Accident research,
driver-assistance systems 1805
Accident situation,
driver-assistance systems 1805
Accident types, integral safety 1426
ACCplus 1894
Accumulator chamber,
antilock braking system 1265
Accumulator volume, common rail 943
Accuracy requirements, sensors 1717
Accuracy, GPS 1873
Accustomization process,
display and control 1791
ACEA specifications, lubricants 545
ACEA standards, lubricants 545
Acid concentration, pH value 200
Acid layer, starter batteries 1484
Acid-base pair 199
Acidic strength, acids in water 198
Acidity constant, acids in water 198
Acids in water 198
Ack Field, CAN 1695
Ackermann angle, chassis systems 1120
Ackermann condition, steering 1193
Acknowledge, vehicle
communication networks 1691
ACME neck, LPG operation 892
Acoustic channel,
display and control 1784
Acoustic glazing,
automotive glazing 1378
Acoustic impedance, acoustics 66
Acoustic quantities, legal units 39
Acoustic tuning devices,
exhaust-gas system 797
Acoustic wind tunnel 528
Acoustics 66
Acoustics, air-intake system 757
Acrylate adhesive,
adhesive technologies 457
Activation arc diode, ignition coil 863
Activation spark suppression,
ignition coil 863
Activation strategy,
night vision systems 1879
Activation,
high-pressure fuel injector 850
Activation,
intake-manifold fuel injector 848
Active clamping, inverters 1585
Active control, ESC 1285
Active mass, starter batteries 1479
Active materials, starter batteries 1480
Active pedestrian protection,
emergency-braking systems 1920
Active power, electrical engineering 127
Active rpm sensors 1738
Active safety functions,
driver-assistance systems 1807
Active safety, safety systems
in the motor vehicle 1412
Active safety, vehicle bodies,
commercial vehicles 1332
Active sound, vehicle acoustics 536
Active spring systems,
suspension 1133
Active star, vehicle
communication networks 1689
Active steering stabilization
on the front axle 1305
Active Steering, steering 1200
Active vehicle chassis systems 1133
Active vibration isolation 63
Active-matrix LCD,
display and control 1789
Activity coefficient,
law of mass action 198
Activity, law of mass action 198
Actuator redundancy concept,
architecture of electronic
systems 1674
Adaptation facility, knock control 859
Adaptation with feedback,
control engineering 247
Adapter circuits,
electronic control unit (ECU) 1632

Adaptive controllers, control engineering 245
Adaptive Cruise Control 1890
Adaptive cruise control, advanced rider-assistance systems for two-wheelers 1930
Adaptive hydraulic absorbers, shock absorbers 1139
Adaptive light distribution pattern, cornering headlamps, lighting equipment 1356
Adaptive pressure control, transmission control 633
Adaptive rear-lighting systems, lighting equipment 1366
Adaptive shock absorber, chassis systems 1142
Adaptive tail lamps, lighting equipment 1367
Additive system, exhaust-gas treatment, diesel engines 964
Additives, diesel fuel 574
Additives, gasoline 567
Additives, lubricants 538
Addressing types, vehicle communication networks 1688
Addressing, MOST 1709
Addressing, Serial Wire Ring 1712
Addressing, vehicle communication networks 1688
Adhesion to road surface, motor-vehicle dynamics 472
Adhesion, tire grip 1178
Adhesion/slip curves, wheel-slip control systems 1260
Adhesive technologies, permanent joints 455
Adhesives, adhesive technologies 455
Adiabatic change of state, thermodynamics 91
Adjacent-channel interference, receivers 1799
Adjustable shock absorbers, chassis systems 1138
Adjustment dimension, lighting equipment 1371
Adjustment, radar sensor 1893
Administration functions, MOST 1709
Adsorber elements, air-intake system 758
Advanced rider-assistance systems for two-wheelers, driver-assistance systems 1930
Advanced Rollover Sensing, occupant-protection systems 1417
Advance-warning time, emergency-braking systems 1917
Aeroacoustics 534
Aeroacoustics, aerodynamics 522
Aerodynamic drag, running resistances 465
Aerodynamic efficiency, aerodynamics 520
Aerodynamic forces, aerodynamics 516
Aerodynamic forces, motor-vehicle dynamics 484
Aerodynamic moments, motor-vehicle dynamics 484
Aerodynamic-drag power, running resistances 465
Aerosol, substances 192
AFS functions, lighting equipment 1356
Agente Reductor Liquido de Óxido de Nitrogénio Automotivo, AdBlue 590
Agility, driving dynamics 1304
Aging stabilizers, gasoline 568
Aging, catalytic converter 876
Aging, plastics 310
AGM battery 1484
AI-based sensor-data fusion 1865
Air belt, occupant-protection systems 1421
Air cleaning, climate control 1393
Air cooling, batteries for electric and hybrid drives 1614
Air cooling, engine cooling 736
Air cooling, reciprocating-piston engine 689
Air deficiency, mixture formation, spark-ignition engine 838
Air drier, brake systems, commercial vehicles 1234
Air duct, air-intake system 755
Air expenditure, charge cycle 664
Air filters, air-intake system 769
Air filtration, air-intake system 755, 770
Air flow, charge cycle 664
Air heaters, auxiliary heater systems 1398
Air inlet, climate control 1392
Air outlet, climate control 1393
Air pollutants, air-intake system 756
Air processing, brake systems, commercial vehicles 1233
Air ratio, mixture formation, spark-ignition engines 838
Air reservoir, brake systems, commercial vehicles 1230
Air spring with bellows, suspension 1129
Air springs, suspension 1126

Air supply, brake systems, commercial vehicles 1233
Air surplus, mixture formation, spark-ignition engines 839
Air System, Motronic 803
Air vents, climate control 1393
Air, refrigerants 592
Airbag, occupant-protection systems 1419
Air-cooled intercoolers 743
Air-directed combustion process, gasoline direct injection 845
Air-filter elements, air-intake system 756
Air-flow sensor, flow meters 1747
Air-gap spark concept, spark plug 870
Air-intake systems 754
Air-mass flow, supercharging 671
Air-processing unit, brake systems, commercial vehicles 1236
Air-water recirculation cooling, reciprocating-piston engine 688
Alarm detectors, theft-alarm system 1448
Alarm signals, theft-alarm system 1448
Alarm siren, theft-alarm system 1448
Alarm systems 1448
Alcohol operation, alternative gasoline-engine operation 900
Aliphatic compounds, fuels 560
Alkaline etching, micromechanics 1720
Alloy materials, starter batteries 1479
Alloy, metallic materials 271
Alloyed high-grade steels, EN metallurgy standards 286
Alloyed stainless steels, EN metallurgy standards 286
Alloying elements, starter batteries 1479
All-pole short-circuiting, inverters 1584
All-wheel drive, drivetrain with internal-combustion engine 598
Alternating-current charging, on-board charger 1600
Alternative A/C-unit concepts 1394
Alternative operating strategies, internal-combustion engines 662
Alternator capacity utilization, 12 V vehicle electrical system 1461
Alternator circuits, alternators 1492
Alternator regulator, 12 V vehicle electrical system 1452
Alternator versions, alternators 1502
Alternators, automotive electrical systems 1490
Altitude correction, Electronic Diesel Control 913
Aluminum alloys, nonferrous metals 277
Aluminum cast wheels 1163
Aluminum forged wheels 1163
Aluminum wheels 1163
Aluminum-sheet wheels 1163
Ammonia, exhaust-gas treatment, diesel engines 967
Amorphous thermoplastics, plastics 307
Amount of substance, substance concentrations 194
Ampere, SI units 33
Ampere-turns, electrical engineering 118
Amperometric measurement, hydrogen sensors 1774
Amplitude modulation, radio and TV broadcasting 1795
Amplitude, vibrations and oscillations 60
AMR effect, electrical engineering 133
AMR sensor elements, electrical engineering 133
AMR sensor, rpm sensors 1739
AMR sensors 1734
Analog interface, electronic control unit (ECU) 1632
Analog-digital converters, electronic control unit (ECU) 1626
Ancillary circuits, brake systems, commercial vehicles 1236
Anechoic chambers, electromagnetic compatibility 1532
Anergy, thermodynamics 92
Angle of incidence, geometric optics 76
Angle of reflection, geometric optics 76
Angle of refraction, geometric optics 76
Angle, legal units 35
Angle/time system, axial-piston distributor pump 950
Angle-of-rotation sensor, axial-piston distributor pump 950
Angles of geometric visibility, lighting equipment 1346
Angular frequency, vibrations and oscillations 60
Angular momentum, mechanics 49
Angular velocity, mechanics 46
Angular-contact ball bearing, rolling bearings 378
Angular-position sensors 1724
Anion, chemical bonds 188

Anions, chemistry 202
Anisotropic etching, micromechanics 1720
Anisotropic magnetoresistive effect, electrical engineering 133
Anisotropic magnetoresistive effect, position sensors 1734
Annealing, heat treatment 335
Annular orifice, piezo high-pressure fuel injector 852
Anodic bonding, micromechanics 1722
Anodic protection, corrosion protection 345
Anodic subprocess, corrosion processes 340
Anodized coatings, deposits and coatings 353
Anodizing process, micromechanics 1722
Anomaly, water 193
Antenna directional characteristic, electrical engineering 130
Antenna factor, electrical engineering 130
Antenna gain, electrical engineering 130
Antenna system, radar-sensor system 1820
Antennas, electrical engineering 129
Anti-aging agents, radial tire 1171
Antiderivative, mathematics 217
Antiderivative, technical statistics 230
Antiferromagnets, magnetic materials 290
Antifoaming agents, diesel fuel 574
Antifreeze, water cooling 737
Anti-friction coatings, lubricants 539
Antilock braking system, wheel-slip control systems 1261
Antimony, starter batteries 1479
Anti-snow system, air-intake system 759
Anti-theft systems, vehicle security systems 1444
Anti-wheelie control, Motorcycle Stability Control 1302
API classification grades, engine oils 546
API classifications, transmission oils 548
Apparent power, electrical engineering 127
Application force, brake 1209
Application forces, mechanics 50
Application software, automotive software engineering 1647, 1648
Application, automotive software engineering 1662
Application, electronic control unit (ECU) 1638
Application, Electronic Diesel Control 915
Application-specific output stages, electronic control unit (ECU) 1634
Approval codes and symbols, lighting equipment 1335
APSM process, micromechanics 1721
Aquaplaning, motor-vehicle dynamics 473
Aquaplaning, tires 1174
Aqueous Urea Solution, AdBlue 590
Arbitration Field, CAN 1695
Arbitration process, Ethernet 1704
Arbitration scheme, CAN 1694
Arbitration, CAN 1695
Arbitration, FlexRay 1699
Arbitration, vehicle communication networks 1686
Arc functions, mathematics 209
Arc, ignition 854
Architecture of electronic systems, automotive electronics 1666
Architecture, electronic control unit (ECU) 1620, 1632
Arcs, lighting equipment 1349
Area, legal units 35
Area-wide information, traffic telematics 1802
Arithmetic mean, technical statistics 227
Arithmetic roughness average, tolerances 357
Arm window lift, power-window systems 1404
Armature-winding system, alternators 1492
Aromatic compounds, fuels 560
Articulated buses, vehicle bodies, commercial vehicles 1330
Articulated connecting rod, reciprocating-piston engine 682
Articulated road train, vehicle bodies, commercial vehicles 1324
Articulation index, acoustics 72
Artificial element transmutation, elements 186
Artificial intelligence, control of a hybrid drive 1016
Artificial intelligence, mathematics and methods 248
Artificial intelligence, sensor-data fusion 1865

Artificial intelligence, video sensors 1855
Artificial neural networks, artificial intelligence 252
Ash, lubricants 538
ASIC chips, electronic control unit (ECU) 1630
Assembled representation, circuit diagrams 1547
Assembled-representation diagram, circuit diagrams 1552
Assembly and connection technology, electronic control unit (ECU) 1636
Assembly injection molding, two-component seals 408
Assembly stress, threaded fasteners 441
Assembly, plug-in connections 1521
Assessment methods, vehicle acoustics 535
Assistant for leaving a parking space, parking and maneuvering systems 1886
Association for Standardization of Automation and Measuring Systems 1647
Asynchronous machine, electrical machines 1572
ATF, lubricants 538
Atmospheric conditions, reciprocating-piston engine 705
Atomic absorption spectroscopy, elements 188
Atomic bonds, chemical bonds 189
Atomic mass, elements 182
Atomic number, elements 182
Atomic orbitals, chemical bonds 190
Atomic structure, electrical engineering 103
Atoms, chemistry 182
Auger electron spectroscopy, elements 187
Auger process, elements 187
Augmented-reality head-up display, display and control 1792
Austempering, heat treatment 334
Austenite, steels 272
Austenitic structure, steels 272
Authentication, vehicle immobilizer 1445
Autodissociation, acids in water 199
Auto-ignition, spark plug 868, 871
Auto-ignitions, gasoline engine 654
Auto-ignitions, ignition 859
Automated manual transmissions, step-up gear units 617
Automated valet parking, future of automated driving 1937
Automated vehicle guidance, driver-assistance systems 1806
Automatic adjusters, drum brake 1258
Automatic brake application, automatic brake-system operations 1295
Automatic brake system 1206
Automatic brake-system operations 1292
Automatic braking, automatic brake-system operations 1295
Automatic emergency braking for vulnerable road users, emergency-braking systems 1920
Automatic emergency braking, emergency-braking systems 1916
Automatic interference suppression, receivers 1800
Automatic leveling control, lighting equipment 1368
Automatic load-dependent braking-force regulator, wheel-slip control systems 1264
Automatic load-sensitive braking-force regulator, brake systems, commercial vehicles 1239
Automatic planetary transmissions, step-up gear units 620
Automatic prefill, automatic brake-system operations 1294
Automatic self-adjusting mechanism, disk brake 1251
Automatic starting systems, starters for passenger cars 1512
Automatic transmission fluids, lubricants 538
Automotive antenna glass 1378
Automotive glazing, vehicle bodies 1374
Automotive insulation glass, automotive glazing 1377
Automotive Open System Architecture 1648
Automotive software engineering 1646
Automotive steering systems, chassis systems 1192
Automotive windshield and window glass, automotive glazing 1375
Autonomous driving, future of automated driving 1934
Autonomous parking, parking and maneuvering systems 1888
AUTOSAR standard, architecture of electronic systems 1677

AUTOSAR 1648
Auxiliary driving lamps, lighting equipment 1343
Auxiliary heater systems, passenger-compartment climate control 1398
Avalanche breakdown, electronics 140
Average peak-to-valley height, tolerances 357
Avogadro's number, substance concentrations 194
Axial play, rolling bearings 379
Axial-piston distributor pumps 948
Axle sensors 1731
Axle steering angle, chassis systems 1115
Axle-load distribution, chassis systems 1114
Axles, axles and shafts 359

B

B6 bridge circuit, alternators 1493
Backup spark, ignition coil 863
Bag Mini Diluter, exhaust-gas measuring techniques 1080
Bag-in-belt system, occupant-protection systems 1421
Bakelite, plastics 319
Balance weights, wheels 1169
Balancing of masses in a multiple-cylinder engine 694
Balancing of masses in a single-cylinder engine 693
Bar, finite-element method 239
Bar, legal units 37
Barber-pole sensor, position sensors 1735
Barber-pole structure, electrical engineering 134
Barium nitrate, NO_x storage catalytic converter 875
Barium oxide, NO_x storage catalytic converter 875
Barium sulfate, NO_x storage catalytic converter 876
Bark scale, acoustics 71
Base constant, bases in water 199
Base Line, alternators 1505
Base metals, chemistry 203
base SI units, quantities and units 32
base SI units, SI units 32
Bases in water 199
Base-speed range, electrical machines 1566
Basic axioms, geometric optics 75
Basic CAN controllers 1696
Basic equations, flow mechanics 58
Basic ignition point, ignition 855
Basic software, automotive software engineering 1648
Basic transverse-dynamics controller, ESC 1277
Batteries for hybrid and electric vehicles, electrics for electric and hybrid drives 1606
Batteries for truck electrical systems, vehicle electrical systems for commercial vehicles 1469
Battery case, starter batteries 1478
Battery designs, starter batteries 1483
Battery discharge power, vehicle electrical systems for hybrid and electric vehicles 1564
Battery installation locations, 12 V vehicle electrical system 1452
Battery maintenance, starter batteries 1488
Battery malfunctions, starter batteries 1488
Battery management system, batteries for electric and hybrid drives 1612
Battery management, 12 V vehicle electrical system 1465
Battery pack, batteries for electric and hybrid drives 1612
Battery parameters, starter batteries 1482
Battery sensor, 12 V vehicle electrical system 1465
Battery system, batteries for electric and hybrid drives 1606
Battery terminal, starter batteries 1478
Battery testers 1488
Battery-status recognition, 12 V vehicle electrical system 1465
Bayonet neck, LPG operation 892
BCD hybrid technology, electronics 149
Bead core, radial tire 1171
Bead, radial tire 1171
Beam, finite-element method 239
Bearing clearance, rolling bearings 379
Bearing forces, reciprocating-piston engine 700
Bearing housing, exhaust-gas turbocharger 776
Bearing materials, friction bearings 373
Bearing play, rolling bearings 379
Bearing surface, reciprocating-piston engine 701
Beat, vibrations and oscillations 61

Becquerel, SI units 34
Beginning of injection period, unit injector 945
Bel, legal units 39
Belt drives, internal-combustion engines 720
Belt pretensioners, occupant-protection systems 1418
Belt pulleys, belt drives 722
Belt run, belt drives 721
Belt slack, occupant-protection systems 1418
Belt-driven starter generator, hybrid drives 995
Belt-force limiter, occupant-protection systems 1419
Belt-pulley profile, belt drives 722
Belt-reminder system, safety systems in the motor vehicle 1410
Belt-run forces, belt drives 721
Belt-run lengths, belt drives 728
Belt-tensioning systems, belt drives 724, 729
Bending stress, material parameters 262
Bending stress, mechanics 52
Bernoulli equation, flow mechanics 58
Bernoulli's Law, flow meters 1746
Bessel functions, wave optics 78
Bevel crown wheel, differentials 626
Bevel-gear differential, differentials 626
Bevel-gear pinion, differentials 626
Bifuel vehicles, engines fueled by natural gas 894
Bifuel-Motronic 805
Bi-Litronic, lighting equipment 1351
Bimorphous spring element, acceleration sensor 1753
Binary classifier, sensor-data fusion 1866
Binary system, mathematics 204
Bingham bodies 538
Biodiesel, diesel fuel 575
Bioethanol, gasoline 568
Biofunctional filters, cabin filters 1397
Biogas, engines fueled by natural gas 894
Biomass-to-Liquid, fuels 577
Biomethane, engines fueled by natural gas 894
Biomethane, natural gas 580
Bipolar transistors, electronics 144
Bit stuffing, CAN 1696
Bitstream, FlexRay 1700
Bi-turbo supercharging, complex supercharging systems 784
Black malleable cast iron, metallic materials 277
Blank, radial tire 1172
Bleeding, lubricants 538
Blend components, diesel fuel 575
Blend components, gasoline 568
Blending chamber, climate control 1393
Blending-flap system, climate control 1393
Blends, plastics 315
Blind rivet, riveting 458
Blind-spot detection, advanced rider-assistance systems for two-wheelers 1931
Block copolymers, plastics 315
Block diagram 1546
Blocking converter, DC-DC converters 1595
Blow-off valve, exhaust-gas turbocharger 779
Bodies submerged in a fluid flow, flow mechanics 59
Body acoustics 534
Body mass, motor-vehicle dynamics 489
Body natural frequency, motor-vehicle dynamics 489
Body shock absorbers, chassis systems 1136
Body springs, suspension 1124
Body structures, vehicle bodies, commercial vehicles 1329
Body types of passenger cars, overall motor-vehicle system 30
Boiling curve, gasoline 565
Boiling range, diesel fuels 570
Boiling range, fuels 562
Boiling temperature, material parameters 258
Bolt connections, positive or form-closed joints 428
Bolt loosening torque, threaded fasteners 441
Bolt tightening torque, threaded fasteners 441
Bolt-through design, reciprocating-piston engine 681
Bond angle, chemical bonds 190
Bond-frame module, inverters 1591
Boost converter, DC-DC converters 1595
Boost pressure, exhaust-gas turbocharging 672
Booster heater, climate control 1393
Booster heaters, auxiliary heater systems 1402

Booster voltage,
high-pressure fuel injector 850
Booster voltage,
solenoid-valve injector 933
Boost-pressure control,
Electronic Diesel Control 913
Boost-pressure control,
exhaust-gas turbocharger 781
Boot, exhaust-gas system 796
Borderline pumping temperature,
lubricants 539
Bore, radial shaft seal 401
Boron treatment, heat treatment 338
Bosch process, DRIE process 1720
Bottleneck/constriction assistant,
lane assistance 1910
Bottom dead center, charge cycle 664
Bottom-feed principle, LPG injector 891
Bottom-up approach, architecture
of electronic systems 1679
Boundary descriptor, MOST 1708
Brake booster,
passenger-car brake system 1217
Brake by wire,
electrohydraulic brake 1225
Brake by wire, mechatronics 1641
Brake circuit, brake systems 1214
Brake coefficient 1209
Brake control,
antilock braking system 1262
Brake coupling head, brake systems,
commercial vehicles 1232
Brake disks, disk brake
for commercial vehicles 1256
Brake disks, disk brake
for passenger cars 1252
Brake equipment 1206
Brake fluids, operating fluids 556
Brake intervention with locking-
differential effect, differentials 629
Brake master cylinder with captive piston
spring,
passenger-car brake system 1220
Brake master cylinder with central valve,
passenger-car brake system 1219
Brake master cylinder,
passenger-car brake system 1219
Brake pads, disk brake 1253
Brake preparation,
emergency-braking systems 1917
Brake slip,
wheel-slip control systems 1260
Brake systems,
definitions and principles 1206
Brake systems, definitions 1206
Brake systems, requirements 1211
Brake systems,
structure and organization 1214
Brake 1207
Brake-assistant operations, automatic
brake-system operations 1292
Brake-circuit configuration,
brake systems 1214
Brake-disk wiper, automatic
brake-system operations 1293
Brake-fluid reservoir,
passenger-car brake system 1221
Brake-pressure modules, electronically
controlled brake system 1243
Brake-slip controller, ESC 1277
Braking deceleration 1210
Braking deceleration, brakes 1210
Braking distance 1210
Braking factor, brake systems 1210
Braking force 1209
Braking in a turn/curve, driving-
dynamics test procedures 503
Braking in a turn/curve,
Motorcycle Stability Control 1299
Braking power 1210
Braking systems for
commercial vehicles 1230
Braking time,
motor-vehicle dynamics 476
Braking torque 1209
Braking work, brake systems 1210
Braking-force distribution 1209, 1215
Braking-time reserve,
lane assistance 1911
Braking-torque compensation,
chassis systems 1118
Brass, nonferrous metals 277
Brayton process, thermodynamics 95
Break points, circuit diagrams 1549
Breakdown angular frequency,
motor-vehicle dynamics 487
Breakdown voltage, electronics 140
Bright chrome,
deposits and coatings 349
Brinell hardness, heat treatment 329
Broad-band interference,
electromagnetic compatibility 1526
Broadcast, vehicle
communication networks 1688
Bronze, nonferrous metals 277
Browning coatings,
deposits and coatings 353
Buck converter, DC-DC converters 1595
Bucket tappet,
gasoline direct injection 830

Buckle tightener, occupant-protection systems 1419
Buffer substance, coolants 554
Bulk Current Injection, electromagnetic compatibility 1531
Bulk micromechanics, yaw-rate sensor 1743
Buoyancy, hydrostatics 57
Burn-off, hot-wire air-mass meters 1748
Bus access, Ethernet 1704
Bus access, FlexRay 1698
Bus access, LIN 1701
Bus arbitration, CAN 1694
Bus configuration, CAN 1694
Bus conflict, vehicle communication networks 1687
Bus driver, vehicle communication networks 1684
Bus states, CAN 1692
Bus systems, vehicle communication networks 1682
Bus topology, vehicle communication networks 1688
Buses in motor vehicles 1692
Buses, vehicle bodies, commercial vehicles 1330
Butane Working Capacity, air-intake system 758
Butt-seam welding, welding 452
Bypass applications, automotive software engineering 1657
Bypass control, exhaust-gas turbocharger 782
Bypass oil filters, engine lubrication 753
Bypass valve, exhaust-gas turbocharging 672
Byte Start Sequence, FlexRay 1700
By-wire mode, brake systems for passenger cars 1227

C

Cab behind engine, vehicle bodies, commercial vehicles 1328
Cabin filters, passenger-compartment climate control 1397
Cable-set model, architecture of electronic systems 1677
Cab-over-engine vehicles, vehicle bodies, commercial vehicles 1328
Cache, electronic control unit (ECU) 1624
CAFE value, EPA legislation 1049
Calculation accuracy, finite-element method 234
Calculation of strength, mechanics 51
Calculation programs, snap-on connections 449
Calibration data, automotive software engineering 1662
Calibration system, automotive software engineering 1662
Calibration tool, automotive software engineering 1662
Calibration, Electronic Diesel Control 915
Caloric equation of state, thermodynamics 89
Calorific value of air/fuel mixture, fuels 562
Cam ring, radial-piston distributor pump 951
Cam shape, variable valve timing 670
Camber angle, chassis systems 1115
Cambering, radial tire 1172
Camera-based interior monitoring, interior-monitoring systems 1932
Cam-follower assembly, reciprocating-piston engine 685
Cam-lobe design, reciprocating-piston engine 687
Camshaft adjustment, variable valve timing 667
Camshaft drive assembly, reciprocating-piston engine 685
Camshaft drive, reciprocating-piston engine 687
Camshaft, reciprocating-piston engine 678
Camshaft-speed sensor 1742
CAN bus, architecture of electronic systems 1667
CAN protocol 1694
CAN, buses in motor vehicles 1692
Candela, SI units 33
CAN-FD, CAN with flexible data rate 1696
Canister-purge valve, fuel supply 824
Cap, seals 396
Capability Maturity Model Integration, automotive software engineering 1652
Capacitance diode, electronics 142
Capacitance, electrical engineering 108
Capacitance, starter batteries 1482
Capacitive coupling, electromagnetic compatibility 1525
Capacitor, electrical engineering 108, 114
Capacity utilization, 12 V vehicle electrical system 1462

Capture function, electronic control unit (ECU) 1626
CARB legislation, emission-control legislation 1043
CARB vehicle categories, overall motor-vehicle system 30
Carbamide, AdBlue 590
Carbon canister, fuel supply 824
Carbon coatings containing metal carbide, deposits and coatings 352
Carbon monoxide, exhaust-gas emissions 717
Carbon-graphite bearings, friction bearings 377
Carbonitriding, heat treatment 337
Carburization depth, heat treatment 337
Carburizing, heat treatment 337
Carnot cycle, thermodynamics 92
Carrier, step-up gear units 610
Cartesian coordinates, mathematics 211
Car-to-car communication, intersection assistant 1926
Car-to-infrastructure communication, intersection assistant 1927
Cascade control, control engineering 243
Case hardening, heat treatment 337
Case-hardening depth, heat treatment 337
Case-hardening steel, metallic materials 276
Casing, radial tire 1171
Cast iron materials, metallic materials 276
Cast resins, plastics 323
Cast steel, metallic materials 277
Caster angle, chassis systems 1119
Caster offset in wheel center, chassis systems 1119
Caster offset, chassis systems 1119
Casting alloys, nonferrous metals 277
Castor offset, Motorcycle Stability Control 1297
Cast-steel engine mount, finite-element method 237
Catalyst layers, fuel cell 1032
Catalyst, exhaust-gas emissions 718
Catalysts, chemical thermodynamics 196
Catalytic converter 792
Catalytic converter, catalytic exhaust-gas treatment 874
Catalytic exhaust-gas treatment, exhaust-gas emissions 718
Catalytic exhaust-gas treatment, management for spark-ignition engines 874
Catalytic measurement principle, hydrogen sensors 1775
Catalytic oxidation, exhaust-gas treatment, diesel engines 962
Catalytic processes, exhaust-gas emissions 719
Catalytic reduction of nitrous fumes, exhaust-gas treatment, diesel engines 965
Catalytic-converter configurations, catalytic exhaust-gas treatment 876
Catalytic-converter diagnosis, on-board diagnostics 1099
Catalytic-converter heating, gasoline direct injection 845
Cathodic counter-reaction, corrosion processes 340
Cathodic deposition, deposits and coatings 350
Cathodic protection, corrosion protection 345
Cation, chemical bonds 188
Cations, chemistry 202
Cavitation, shock absorbers 1137
Cell cyclone, air-intake system 769
Cell voltage, starter batteries 1480
Cellulose-based media, air-intake system 757
Cementite, metallic materials 271
Center electrode, spark plug 867
Center of gravity, chassis systems 1114
Center-hole diameter, wheels 1156
Central display unit, display and control 1788
Central element, ESC 1283
Central introduction point, air-intake system 766
Central locking, locking systems 1433
Central monitor, display and control 1788
Central operating unit, display and control 1788
Central valve, passenger-car brake system 1220
Central ventilation, starter batteries 1485
Centralized E/E architecture, architecture of electronic systems 1670
Centrifugal force, mechanics 47
Centrifugal pendulum-type absorber, vibration isolation 606
Centripetal acceleration, vehicle motion 463

Centroidal distance, mechanics 46
Ceramic materials 303
Ceramic materials, properties 304
Ceramic monolith,
catalytic converter , 792
Ceramics, materials 303
Ceroxide,
three-way catalytic converter 875
Cetane improvers, diesel fuel 574
Cetane index, diesel fuel 570
Cetane number, diesel fuel 569
CFV method, exhaust-gas
measuring techniques 1080
Chain drives, internal-combustion
engines 730
Chain pitch, chain drives 731
Chain tensioner, chain drives 734
Challenge-response scheme,
vehicle immobilizer 1445
Change of linear momentum,
mechanics 49
Changeover valve, ESC 1285
Changes of state,
thermodynamics 86, 91
Characteristic alternator values,
alternators 1498
Characteristic curves,
electrical machines 1571
Characteristic impedance,
electrical engineering 128
Characteristic power curve,
electrical machines 1567
Characteristic running-resistance curve,
drivetrain with internal-combustion
engine 597
Characteristic torque curve,
electrical machines 1567
Characteristic vehicle speed,
ESC 1288
Characteristics, chassis systems 1113
Characteristics, fuels 562
Charcoal filters, cabin filters 1397
Charge balance,
12 V vehicle electrical system 1452
Charge carriers,
electrical engineering 103
Charge cycle 664
Charge cycle, cylinder charge 814
Charge cycle, gasoline engine 652
Charge indicator lamp,
12 V vehicle electrical system 1456
Charge movement,
gasoline engine 653
Charge quantity,
electrical engineering 104
Charge transfer,
electrical engineering 105
Charge, electrical engineering 117
Charge-air pressure sensor 1758
Charge-air temperature sensor 1763
Charge-balance calculation,
12 V vehicle electrical system 1457
Charge-depleting test,
energy demand for drives 509
Charger, on-board charger 1600
Charge-sustaining test,
energy demand for drives 509
Charging and discharging,
starter batteries 1480
Charging characteristic,
starter batteries 1488
Charging infrastructure,
electrification of drive 982
Charging method,
starter batteries 1487
Charging power, on-board charger 1602
Charging strategy, vehicle
electrical systems for hybrid
and electric vehicles 1564
Charging time, ignition coil 862
Charging, starter batteries 1487
Chassis acoustics 534
Chassis configuration,
chassis systems 1110
Chassis dynamometer, exhaust-gas
measuring techniques 1079
Chassis frame, vehicle bodies,
commercial vehicles 1327
Chassis structure, vehicle bodies,
commercial vehicles 1331
Chassis systems 1108
Chassis, vehicle bodies,
commercial vehicles 1325
Chauffeur brake,
electrohydraulic brake 1224
Chemical bonds 188
Chemical copper,
deposits and coatings 350
Chemical equilibrium 197
Chemical nickel,
deposits and coatings 350
Chemical reactions,
starter batteries 1480
Chemical thermodynamics 195
Chemical tin,
deposits and coatings 350
Chemical vapor deposition,
deposits and coatings 352
Chemically deposited coatings,
deposits and coatings 349

Chemiluminescence detector, exhaust-gas measuring techniques 1082
Chemistry 182
Child Occupant Protection, safety systems in the motor vehicle 1410
Child seats, safety systems in the motor vehicle 1410
Child-lock function, locking systems 1433
Chinese legislation, emission legislation for motorcycles 1075
Chinese legislation, emission-control legislation 1056
Chirp sequence, radar-sensor system 1820
Chloroprene rubber, elastomer materials 415
Choke limit, exhaust-gas turbocharger 778
Chrome deposits, deposits and coatings 349
Circuit diagrams 1545
Circuit symbols 1538
Circuit-board technology, electronic control unit (ECU) 1636
Circular keyed joint, frictional joints 436
Circular polarization, wave optics 77
Circular spline, steering 1201
Circular-course driving, steady-state, driving-dynamics test procedures 499
CISC processor, electronic control unit (ECU) 1624
City buses, vehicle bodies, commercial vehicles 1330
Clamp joints, frictional joints 435
Clamping sleeve, frictional joints 434
Class label, sensor-data fusion 1866
Classic riveting, permanent joints 457
Classic states of aggregation, substances 192
Classification of motor vehicles, overall motor-vehicle system 29
Classification, hybrid drives 994
Classifiers, sensor-data fusion 1866
Clausius-Rankine cycle, thermodynamics 95
Claw-pole principle, alternators 1492
Clean-air line, air-intake system 755, 770
Clearance compensation, disk brake 1251
Clearance examination, molded seal 393
Clearance fit, rolling bearings 380
Clearance lamps, lighting equipment 1361
Clearance, disk brake 1251
Climate control for hybrid and electric vehicles 1396
Climate-controlled wind tunnels 528
Climbing power, running resistances 467
Climbing resistance, running resistances 467
Clinch bolt, riveting 458
Clinching, penetration-clinching processes 459
Clock data recovery, Serial Wire Ring 1711
Clock frequency, inverters 1583
Clock rate, Serial Wire Ring 1712
Close range, driver-assistance systems 1809
Closed loop, driving-dynamics test procedures 498
Closed system, thermodynamics 86
Closed vessels, hydrostatics 56
Closed-loop control, control engineering 242
Closed-loop control-circuit influences, inverters 1589
Close-joint soldering, soldering 455
Closing curve, driving-dynamics test procedures 503
Closing times, reciprocating-piston engine 686
Cloud point, lubricants 538
Cloud, ACC ongoing development 1895
Clutch disk, power take-up elements 600
Clutch lining, power take-up elements 600
Clutch torque 800
CMOS transistors, electronics 148
CO_2 emissions, EU legislation 1053
CO_2 emissions, emission legislation for motorcycles 1074
CO_2 lasers, laser welding 453
CO_2, refrigerants 592
Coal-to-Liquid, fuels 577
Coated glazing, automotive glazing 1376
Coatings, corrosion protection 346
Coatings, deposits and coatings 348
Coaxial e-axle topology, electric final-drive unit 1023
COD process, fuels 578
Coefficient of cornering resistance, running resistances 465

Coefficient of friction, mechanic 53
Coefficient of longitudinal expansion, material parameters 258
Coefficient of resistance, flow mechanics 59
Coefficient of rolling friction, mechanics 55
Coefficient of rolling resistance, running resistances 464
Coefficient of thermal expansion, material parameters 258
Coefficient of volume expansion, material parameters 258
Coercive field strength, electrical engineering 122
Coil ignition, ignition 860
Coil in the alternating-current circuit, electrical engineering 116
Coil, electrical engineering 109, 115
Co-injection technique, two-component seals 408
Coke carbons, batteries for electric and hybrid drives 1609
Cold spark plug 868
Cold viscosity, brake fluids 557
Cold-running support, flame start system 960
Cold-start phase, diesel engine 660
Cold-start processes, diesel engine 660
Cold-starting reliability, lubricants 539
Cold-work steels, metallic materials 272
Collision ionization, electronics 140
Collision risk, intersection assistant 1924
Colloid, substances 192
Color filters, geometric optics 77
Column, vehicle communication networks 1686
Comb structure, yaw-rate sensor 1744
Combination brake cylinder, brake systems, commercial vehicles 1240
Combination supercharging, complex supercharging systems 787
Combination systems, common rail 925
Combined insulation materials, plastics 323
Combiner, display and control 1791
Comb-shaped electrodes, acceleration sensor 1754
Combustion bomb, ignition 854
Combustion cycle, charge cycle 665
Combustion in diesel engines, mixture formation 658
Combustion in gasoline engines, mixture formation 649
Combustion knock, spark plug 871
Combustion of homogeneous mixtures, gasoline engine 650
Combustion of partially homogeneous mixtures, gasoline engine 652
Combustion pressure, ignition 855
Combustion processes, gasoline engine 649
Combustion time, ignition 854
Combustion torque, spark-ignition engine 800
Combustion velocity, gasoline engine 651
Comfort and convenience systems, door and roof areas 1404
Comfort, aerodynamics 522
Commercial vehicle battery, automotive electrical systems 1485
Commercial vehicles, overall motor-vehicle system 29
Commercial vehicles, vehicle bodies, commercial vehicles 1324
Commercial-vehicle air filters, air-intake system 770
Commercial-vehicle air-intake system, air-intake system 768
Commercial-vehicle exhaust-gas systems 798
Commercial-vehicle intake manifolds, air-intake system 771
Commercial-vehicle tires 1173
Commercial-vehicle transmissions, step-up gear unit 624
Common rail 924
Common-mode input resistance, operational amplifier 151
Common-mode rejection ratio, operational amplifier 151
Common-rail injection system 924
Common-rail system 924
Communication areas, display and control 1785
Communication controller, vehicle communication networks 1684
Communication interfaces, electronic control unit (ECU) 1635
Communication modes, PSI5 1705
Communication zone, locking systems 1441
Communication, electronic control unit (ECU) 1630
Communication, Motronic 804
Commutation, inverters 1584

Compact alternator, alternators 1503
Compact coil, ignition coil 863
Compact spare wheels 1167
Compact-diode alternator, alternators 1504
Compacting, plastics 322
Compare function, electronic control unit (ECU) 1626
Comparison cycles, thermodynamics 92
Compensating chamber, shock absorbers 1137
Compensating springs, suspension 1131
Complete balancing of masses, reciprocating-piston engine 694
Completely maintenance-free batteries, starter batteries 1483
Complex numbers, mathematics 210
Complex supercharging systems, turbochargers and superchargers 784
Component forces, aerodynamics 523
Component measurement methods, electromagnetic compatibility 1530
Component model, architecture of electronic systems 1676
Component protection 838
Component protection, catalytic converter 880
Component safety, frictional joints 433
Composability, vehicle communication networks 1687
Composability, vehicle communication systems 1687
Composite bearings, friction bearings 376
Composite injection molding, two-component seals 408
Composite materials 303
Composite materials, air-intake system 757
Composite techniques, two-component seals 409
Compound electrodes, spark plug 867
Comprehensive components, on-board diagnostics 1100
Compressed natural gas, engines fueled by natural gas 894
Compressed natural gas, natural gas 580
Compressibility, brake fluids 556
Compression cycle, charge cycle 665
Compression function, technical statistics 229
Compression ignition in gasoline operation, mixture formation 662
Compression movement, wheel suspensions 1147
Compression ratio, reciprocating-piston engine 706
Compression ratio, thermodynamics 93
Compression set, elastomer materials 418
Compression springs, springs 367
Compression stage, shock absorbers 1141
Compression volume, reciprocating-piston engine 706
Compression, O-ring 388
Compressive stress, mechanics 52
Compressive yield point, material parameters 262
Compressive-stress relaxation, elastomer materials 418
Compressor, brake systems, commercial vehicles 1233
Compressor, climate control 1395
Compressor, exhaust-gas turbocharger 777
Computer core, electronic control unit (ECU) 1624
Computer-aided design 234
Computer-aided lighting, lighting equipment 1347
Concave lens, geometric optics 76
Concentrated load, rolling bearings 381
Concentration sensors 1766
Concentricity, wheels 1168
Condensed water, climate control 1393
Condenser, climate control 1395
Conductance, electrical engineering 106
Conducted disturbances, electromagnetic compatibility 1530
Conductive charging concepts, on-board charger 1600
Conductive coupling, electromagnetic compatibility 1525
Conductive glass seal, spark plug 866
Conductive grid, starter batteries 1479
Conductivity, electronics 138
Conductors, electrical engineering 109
Confidence interval, technical statistics 231
Confidence level, technical statistics 231
Conflict diagram, chassis systems 1112
Conformability, friction bearings 372
Conical seal seat, spark plug 867
Connecting elements, exhaust-gas system 796

Connecting rod,
reciprocating-piston engine 677, 679
Connecting-rod eyes,
reciprocating-piston engine 679
Connecting-rod force,
reciprocating-piston engine 692
Connecting-rod mass,
reciprocating-piston engine 693
Connecting-rod violin,
reciprocating-piston engine 679
Connection master, MOST 1709
Consent Decree,
emission-control legislation
for heavy commercial vehicles 1063
Conservation of energy, mechanics 49
Conservation of momentum,
mechanics 49
Consistency, lubricants 539
Constant throttle, continuous-operation
brake systems 1246
Constant-pressure cycle,
thermodynamics 94
Constant-pressure supercharging,
exhaust-gas turbocharger 780
Constant-volume cycle,
thermodynamics 94
Constituents of engine exhaust gas,
exhaust-gas emissions 716
Construction,
electric final-drive unit 1021
Constructive interference, wave optics 78
Consumer classification,
12 V vehicle electrical system 1451
Consumer tests, safety systems
in the motor vehicle 1414
Contact corrosion,
types of corrosion 341
Contact patch, tires 1176
Contact potential, electrical
engineering 131
Contact stress, threaded fasteners 443
Contact systems,
plug-in connections 1523
Contact thermometer 1760
Contact-corrosion current measurement,
corrosion testing 343
Content-based addressing, CAN 1694
Continuity equation, flow mechanics 58
Continuous brake system 1208
Continuous distributions,
technical statistics 229
Continuous loads,
12 V vehicle electrical system 1451
Continuous operation,
electrical machines 1567
Continuous performance,
electric final-drive unit 1026
Continuous λ control,
exhaust-gas treatment 880
Continuous-delivery high-pressure pump,
gasoline direct injection 832
Continuous-delivery system,
fuel supply 823
Continuously variable transmissions,
step-up gear unit 622
Continuous-operation brake
system 1206
Continuous-operation brake system,
brake systems, commercial
vehicles 1233, 1246
Contraction coefficient,
flow mechanics 58
Contrast, lighting quantities 82
Control algorithms, ACC 1892
Control bit, CAN 1695
Control channel, MOST 1707
Control collar,
axial-piston distributor pump 949
Control collar,
radial-piston distributor pump 952
Control cycle,
antilock braking system 1261
Control device, brake systems 1206
Control device,
passenger-car brake system 1217
Control engineering 242
Control Field, CAN 1695
Control filament,
preheating systems 956
Control force, brake 1209
Control messages, MOST 1708
Control modules, ACC 1892
Control of a hybrid drive,
electrification of drive 1010
Control of electric drives,
electrification of drive 984
Control of shift quality,
transmission control 633
Control on the high-pressure side,
common rail 926
Control process,
antilock braking system 1261
Control rod, natural-gas pressure-control
module 899
Control strategy, inverters 1588
Control structure, ESC 1275
Control task, control engineering 244
Control valve, steering 1196
Control-based operating strategy,
control of a hybrid drive 1012

Control-blade independent wheel suspension, wheel suspensions 1152
Controlled braking, automatic brake-system operations 1294
Controlled system, control engineering 242, 245
Controlled variable, control engineering 242
Controller Area Network, buses in motor vehicles 1692
Controller design, control engineering 245
Controllo Stabilità e Trazione, driving-dynamics control system 1287
Convection, thermodynamics 96
Convective heat transfer, thermodynamics 96, 99
Convenience lamps, lighting equipment 1367
Conventional batteries, starter batteries 1483
Conventional radio tuner 1797
Conventional wiper blade, windshield, rear-window and headlamp cleaning systems 1386
Convergent lens, geometric optics 76
Conversion coatings, deposits and coatings 353
Conversion range, hydrodynamic torque converter 604
Conversion rates, catalytic converter 879
Conversion, catalytic converter 874
Converter increase, hydrodynamic torque converter 604
Converter lockup, transmission control 633
Converters, inverters 1582
Convex lens, geometric optics 76
Coolant circuit, reciprocating-piston engine 689
Coolant cooling, water cooling 737
Coolant expansion tank, water cooling 739
Coolant heaters, auxiliary heater systems 1399
Coolant temperature, reciprocating-piston engine 689
Coolant, reciprocating-piston engine 688
Coolant, water cooling 737
Coolant-cooled intercoolers 744
Coolants for internal-combustion engines 553
Coolant-temperature regulation, water cooling 741
Coolant-temperature sensor 1763
Cooler types, inverters 1592
Cooling circuit, exhaust-gas cooling 745
Cooling circuit, water cooling 737
Cooling module, engine cooling 747
Cooling, aerodynamics 521
Cooling, electric final-drive unit 1021
Cooling, electrical machines 1580
Cooling, electronic control unit (ECU) 1636
Cooling, inverters 1591
Cooling, reciprocating-piston engine 688
Cooling-air fan, water cooling 739
Cooling-down phase, flame start system 960
Cooling-system technology, engine cooling 748
Cooperatively regenerative braking system, electrification of drive 976
Coordinate systems, mathematics 211
Coordination Engine, Motronic 804
Copolymers, plastics 307
Copper alloys, nonferrous metals 277
Copper-nickel alloys, nonferrous metals 277
Cords, radial tire 1171
Core-back technique, two-component seals 410
Coriolis acceleration 1743
Coriolis force, yaw-rate sensor 1744
Cornering headlamps, lighting equipment 1355, 1356
Cornering lamps, lighting equipment 1363
Cornering resistance, running resistances 465
Cornering stiffness, motor-vehicle dynamics 484
Correcting variable, control engineering 242
Correction functions, common rail 928
Corresponding acid-base pairs 199
Corrosion inhibitors, diesel fuel 574
Corrosion inhibitors, gasoline 568
Corrosion processes, corrosion and corrosion protection 340
Corrosion protection, brake fluids 558
Corrosion protection, corrosion and corrosion protection 344
Corrosion protection, deposits and coatings 348
Corrosion testing, corrosion and corrosion protection 342

Corrosion, corrosion and corrosion protection 340
Corrosive attack, corrosion processes 340
Corrugated pipe, exhaust-gas system 796
Corrugated tubing, wiring harnesses 1520
Cosine function, mathematics 208
Cotangent function, mathematics 209
Coulomb force, electrical engineering 125
Coulomb friction, mechanics 53
Coulomb, electrical engineering 104
Coulomb, SI units 34
Coulomb's law of sliding friction, mechanics 53
Counterflow cylinder head, reciprocating-piston engine 683
Countershaft, step-up gear units 615
Countersunk-head rivet, riveting 458
Coupler, hydraulic coupler 935
Coupler, piezo high-pressure fuel injector 852
Couplers, electrical engineering 129
Coupling lever, reciprocating-piston engine 682
Coupling, vibrations and oscillations 61
Coupling-shaft link, wheel suspensions 1148
Course setting, ACC 1891
Covalent bonds, chemical bonds 189
Cracking, reciprocating-piston engine 679
Cradle to grave, energy demand for drives 514
Crank pin, reciprocating-piston engine 680
Crankcase scavenging, 2-stroke cycle 666
Crankcase ventilation, air-intake system 760
Crankcase, reciprocating-piston engine 676, 681
Crankcase-ventilation gases, air-intake system 766
Crankshaft assembly, reciprocating-piston engine 691
Crankshaft drive, reciprocating-piston engine 676
Crankshaft main bearing 680
Crankshaft starter generator, hybrid drives 995
Crankshaft throw, reciprocating-piston engine 681
Crankshaft vibration damper 689
Crankshaft web 693
Crankshaft, reciprocating-piston engine 677, 680
Crankshaft-assembly design, reciprocating-piston engine 691
Crankshaft-assembly kinetics, reciprocating-piston engine 692
Crankshaft-drive kinematics, reciprocating-piston engine 691
Crankshaft-speed sensor 1741
Crash recorder, safety systems in the motor vehicle 1411
Crash, safety systems in the motor vehicle 1411
CRC Field, CAN 1695
CRC sequence, CAN 1696
Creaking, locking systems 1434, 1437
Creep behavior, plastics 309
Creeping function, control of electric drives 986
Crevice corrosion, types of corrosion 341
Crimping process, plug-in connections 1523
Critical driving situations, driver-assistance systems 1804
Critical point, water 193
Critical speeds in curves, motor-vehicle dynamics 485
Critical stress intensity factor, material parameters 262
Criticality measurements, driver-assistance systems 1917
Cross product, coriolis acceleration 1743
Cross product, mathematics 214
Crossflow cylinder head, reciprocating-piston engine 683
Cross-linking, adhesive technologies 455
Cross-member, vehicle bodies, commercial vehicles 1328
Cross-over body shapes, vehicle bodies, passenger cars 1313
Cross-ply tires 1171
Cross-section, O-ring 386
Cross-traffic assistant, intersection assistant 1926
Cruise control, ACC 1891, 1892
Cruise control, Electronic Diesel Control 913
Crystal structure, electrical steel strip 292
CTS system, wheels 1158, 1167

Cubic coefficient of thermal expansion, material parameters 258
Cumulative shield, inverters 1585
Cup, radial shaft seal 399
Curie point, material parameters 259
Curie temperature, magnetic materials 290
Curie temperature, material parameters 259
Current characteristic, alternators 1498
Current density, electrical engineering 105
Current intensity, electrical engineering 104
Current law, electrical engineering 110
Current output capability, starter batteries 1483
Current, electrical engineering 104
Cushioning seal, pneumatic seals 406
Customer code area, motor sport 884
Customer-specific extra-small series, motor sport 885
Cutter FID, exhaust-gas measuring techniques 1083
CVD coatings, deposits and coatings 352
CVO process, high-pressure fuel injector 850
CVS dilution procedure, exhaust-gas measuring techniques 1079
CW phase measurement method, lidar-sensor system 1833
Cycle count, FlexRay 1699
Cycle direction, thermal engines 643
Cycle stability, 12 V vehicle electrical system 1459
Cycle stability, starter batteries 1483
Cycles, thermodynamics 91
Cycles, vehicle communication networks 1686
Cyclic redundancy check, Serial Wire Ring 1714
Cyclic redundancy check, vehicle communication networks 1691
Cyclic variations, gasoline engine 653
Cyclical variation, vehicle electrical systems for hybrid and electric vehicles 1564
Cyclone principle, air-intake system 759
Cyclone, air-intake system 769
Cylinder arrangements, reciprocating-piston engine 690
Cylinder charge, management for spark-ignition engines 812
Cylinder head, reciprocating-piston engine 677, 683
Cylinder shutoff, cylinder charge 817
Cylinder-bank shutoff, cylinder shutoff 817
Cylinder-charge movement, diesel engine 659
Cylinder-head gasket, reciprocating-piston engine 684
Cylinder-head gaskets, flat seals 395
Cylinder-head identification, reciprocating-piston engine 683
Cylindrical coordinates, mathematics 212
Cylindrical interference fit, frictional joints 430
Cylindrical rolling bearing, rolling bearings 378

D

Daisy-chain topology, vehicle communication networks 1690
Damped natural frequency, motor-vehicle dynamics 488
Dampers, shock absorbers 1137
Damper-strut independent wheel suspension, wheel suspensions 1152
Damping characteristic curve, shock absorbers 1141
Damping characteristics, shock absorbers 1140
Damping coefficient, acceleration sensor 1750
Damping constant, shock absorbers 1140
Damping control, shock absorbers 1141
Damping factor, motor-vehicle dynamics 482, 489
Damping force, acceleration sensor 1750
Damping force, shock absorbers 1140
Damping functions, antilock braking system 1265
Damping ratio, vibrations and oscillations 61
Damping, vibrations and oscillations 61
Danger of jackknifing, ESC, commercial vehicles 1288
Danger of overturning, ESC 1281
Darlington transistor, electronics 145
Data Field, CAN 1695
Data field, LIN 1702
Data field, Serial Wire Ring 1713

Data frame counter, Serial Wire Ring 1713
Data frame, CAN 1695
Data frame, Ethernet 1704
Data frame, FlexRay 1699
Data frame, LIN 1701
Data frame, Serial Wire Ring 1712
Data frames, MOST 1708
Data matrix code, piezo high-pressure fuel injector 852
Data memory, electronic control unit (ECU) 1629
Data memory, vehicle navigation 1875
Data poisoning, artificial intelligence 256
Data processing, electronic control unit (ECU) 1632
Data processing, Electronic Diesel Control 910
Data transfer, MOST 1708
Data transmission, Serial Wire Ring 1712
Data, vehicle communication networks 1682
Data_0, FlexRay 1697
Data_1, FlexRay 1697
Data-link layer, vehicle communication networks 1685
Data-transfer rate, PSI5 1705
Daytime running lamps, lighting equipment 1362
Daytime vision, lighting quantities 81
DC-DC converters of high voltages to 12 V 1598
DC-DC converters of high voltages to 48 V 1599
DC-DC converters, electrics for electric and hybrid drives 1594
Dead lock, locking systems 1438
Dead reckoning, vehicle navigation 1873
Decibel 67
Decibel, legal units 39
Decimal system, mathematics 204
Decompression brake, continuous-operation brake systems 1246
Decoupling element, exhaust-gas system 796
Dedicated hybrid transmissions, hybrid drives 1008
Deep etching, micromechanics 1721
Deep learning, artificial intelligence 253
Deep neural networks, sensor-data fusion 1867
Deep-bed filter media, engine lubrication 752
Deep-bed filters, air-intake system 756
Deep-cycle resistant battery, starter batteries 1485
Deep-groove ball bearing, rolling bearings 378
Deep-welding effect, welding 453
Definition of motor vehicle, overall motor-vehicle system 28
Definitions of power, reciprocating-piston engine 705
Deflection-force lever arm, chassis systems 1119
Defoamers, diesel fuel 574
Deformation calculation, finite-element method 234
Deformation energy hypothesis, mechanics 51
Defrost vent, climate control 1394
Degassing, starter batteries 1478
Degree Celsius, SI units 34
Degree of reflection, geometric optics 76
Delivery module, exhaust-gas treatment, diesel engines 968
Delivery rate, high-pressure pump, common rail 939
Delivery stroke, unit injector 945
Delta connection, alternators 1492
Demagnetization curves, electrical engineering 123
Demand control, high-pressure pump, common rail 939
Demand-controlled system, fuel supply 822
Demand-regulated high-pressure pump, gasoline direct injection 830
Demand-regulated returnless system, fuel supply 821
Demand-regulated system, fuel supply 823
Demeshing, starters for passenger cars 1511
Density, elastomer materials 418
Density, material parameters 258
Depleting test, energy demand for drives 509
Depletion type, field-effect transistor 147
Deposits, deposits and coatings 348
Depowered airbag, occupant-protection systems 1420
Derating concept, inverters 1588
Derating curve, wiring harnesses 1519
Derating function, inverters 1587
Derivative, mathematics 215

Derived SI units, SI units 34
Descriptive statistics, technical statistics 226
Desiccant box, brake systems, commercial vehicles 1234
Design criteria, control engineering 245
Design fields of vehicle conception, vehicle bodies, passenger cars 1312
Design guidelines, snap-on connections 448
Design of software functions, automotive software engineering 1659
Designation system for aluminum, EN metallurgy standards 289
Designation system for cast iron, EN metallurgy standards 288
Designation system for steels, EN metallurgy standards 286
Desired deceleration input, ESC 1275
Desired drive torque, ESC 1275
Desmotronic, reciprocating-piston engine 687
Destination designation, circuit diagrams 1549
Destination entry, vehicle navigation 1874
Destructive interference, wave optics 78
Desulfating, NO_x storage catalytic converter 876
Detachable connections, joining and bonding techniques 426
Detached representation, circuit diagrams 1548
Detection characteristics, ultrasonic sensor system 1814
Detergents 538
Detergents, diesel fuel 574
Detergents, gasoline 568
Determinant, Mathematics 223
Determining the angle, radar-sensor system 1820
Determinism, FlexRay 1698
Determinism, vehicle communication networks 1683
Development aims, tires 1187
Development methodology, mechatronics 1642
Development of assisting driving functions, future of automated driving 1936
Development of complexity, electronic control unit (ECU) 1622
Development process, architecture of electronic systems 1678
Development process, automotive software engineering 1650
Development tools, architecture of electronic systems 1679
Development work on vehicle acoustics 532
Deviation in concentricity, radial shaft seal 402
Dezincification, types of corrosion 341
Diagnostic Fault Path Management, on-board diagnostics 1098
Diagnostic frequency, on-board diagnostics 1093
Diagnostic Function Scheduler, on-board diagnostics 1098
Diagnostic standards, automotive software engineering 1650
Diagnostic System Management, on-board diagnostics 1098
Diagnostic System, Motronic 804
Diagnostic Validator, on-board diagnostics 1098
Diagnostics 1090
Diagonal-link independent wheel suspension, wheel suspensions 1150
Diamagnetic materials, electrical engineering 122
Diamagnets, magnetic materials 290
Diamond-like carbon coatings, deposits and coatings 352
Diamond-like carbon, deposits and coatings 352
Diaphragm-type pressure regulator, pressure-control module 899
Diaphragm-type sensors, pressure sensors 1756
Dicyclohexylamin nitrite, corrosion protection 346
Die plate, punch riveting 458
Dielectric constant, electrical engineering 118
Dielectric, electrical engineering 117
Diesel cycle, thermodynamics 94
Diesel distributor injection pumps 948
Diesel engine, mixture formation 657
Diesel Exhaust Fluid, AdBlue 590
Diesel filter 922
Diesel fuel, fuels 569
Diesel particulate filter 793
Diesel smoke-emission test, exhaust-gas measuring techniques 1085
Diesel-engine combustion processes, mixture formation 657
Differential bevel gear, differentials 626
Differential calculus, mathematics 215

Differential equation,
oscillatory system 62
Differential equations, mathematics 218
Differential Hall sensor,
rpm sensors 1739
Differential input resistance,
operational amplifier 151
Differential-lock function,
traction control system 1264
Differential-pressure sensor 1758
Differentials, drivetrain 626
Diffraction pattern, wave optics 78
Diffraction, wave optics 78
Diffusion barrier, λ sensors 1769
Diffusion coatings,
deposits and coatings 353
Diffusion combustion,
diesel engine 659
Diffusion detection, tire-pressure
monitoring systems 1188
Diffusion rate, intercooler 744
Digital control, control engineering 245
Digital Directional Antenna,
receivers 1800
Digital displays, display and control 1787
Digital Diversity System, receivers 1800
Digital equalizer, receivers 1798
Digital exterior-mirror system, information
and warning systems 1898
Digital instrument cluster, infotainment
and cockpit solutions 1782
Digital interface,
electronic control unit (ECU) 1632
Digital key, locking systems 1441
Digital map, vehicle navigation 1875
Digital modulation processes,
radio and TV broadcasting 1795
Digital receiver 1798
Digital Sound Adjustment,
receivers 1799
Dilatant flow behavior, lubricants 542
Dilution system, exhaust-gas
measuring techniques 1079
Dilution systems, exhaust-gas
measuring techniques 1080
Dimensional deviations, tolerances 354
Dimensional tolerances, tolerances 355
Dimethyl ether, fuels 584
DI-Motronic 805
DI-Motronic,
homogeneous operation 807
DI-Motronic, lean-burn concept 808
Diodes, electronics 141
Dipole antennas,
electrical engineering 129
Dipole moment,
electrical engineering 117
Dipole-dipole interactions,
chemical bonds 191
Dirac delta function, mathematics 221
Direct control, transmission control 640
Direct LPG liquid injection,
LPG operation 889
Direct Memory Access,
electronic control unit (ECU) 1626
Direct pulse time-of-flight method,
lidar-sensor system 1829
Direct-current charging,
on-board charger 1601
Directional diagram,
electrical engineering 130
Directional stability, aerodynamics 520
Directional stability, ESC 1273
Directional tread, tires 1182
Directly measuring tire-pressure
monitoring systems 1188
Direct-shift transmissions,
step-up gear units 618
Dirt sensor 1778
Dirt wipers, pneumatic seals 406
Disability glare,
lighting equipment 1346
Disc springs, springs 365
Discharge coefficient,
flow mechanics 58
Discharge current,
starter batteries 1481
Discharge output, vehicle electrical
systems for hybrid and electric
vehicles 1564
Discharging,
starter batteries 1481, 1487
Discomfort glare,
lighting equipment 1346
Disconnect unit,
electric final-drive unit 1029
Discrete distributions,
technical statistics 228
Discrete semiconductor devices,
electronics 141
Discrete-time control,
control engineering 245
DISH neck, LPG operation 892
Dish recess,
reciprocating-piston engine 676
Disk brakes for commercial vehicles,
wheel brakes 1255
Disk brakes for passenger cars,
wheel brakes 1250
Disk oil coolers, oil cooling 746

Disk wheel, wheels 1154
Disk-stack oil coolers, oil cooling 746
Dispersants 538
Dispersion parameters, technical statistics 227
Dispersion, geometric optics 75
Dispersion, substances 192
Displacement density, electrical engineering 117
Displacement flux, electrical engineering 117
Displacement, reciprocating-piston engine 706
Display and control, infotainment 1784
Display types, display and control 1789
Display unit, display and control 1788
Dissociation, law of mass action 198
Distance accuracy, lidar-sensor system 1827
Distance calculation, ultrasonic sensor system 1815
Distance measurement, lidar-sensor system 1824
Distance measurement, radar-sensor system 1819
Distance measurement, ultrasonic sensor system 1813
Distillation, fuels 561
Disturbance, control engineering 242
Divergent lens, geometric optics 76
DIY and dealer-installed fog lamps, lighting equipment 1353
DLC coatings, deposits and coatings 352
Domain, architecture of electronic systems 1668
Domain-centralized E/E architecture, architecture of electronic systems 1668
Domains, mechatronics 1644
Dominant state, CAN 1692
Doped lubricants 539
Doping, electronics 138
Doppler effect, acoustics 67
Doppler effect, GPS 1873
Double bonds, chemical bonds 189
Double lock, locking systems 1438
Double ring, vehicle communication networks 1690
Double-cone synchromesh, step-up gear units 612
Double-disk clutch, vibration isolation 601
Double-ECU concept, LPG operation 890
Double-ECU concepts, engines fueled by natural gas 896
Double-layer capacitors, supercapacitors 1617
Double-shaft transmissions, step-up gear units 615
Double-tube oil coolers, oil cooling 746
Double-wishbone independent wheel suspension, wheel suspensions 1150
Downgrade force, running resistances 467
Downsizing, hybrid drives 997
Draft prevention, aerodynamics 523
Drag area, aerodynamics 519
Drag coefficient, aerodynamics 516
Drag coefficient, flow mechanics 59
Drag coefficient, running resistances 465
Drag power, continuous-operation brake systems 1246
Drag torque, hybrid drives 1000
Drag, aerodynamics 516
Drag-torque simulation, electrification of drive 976
Draw tempering, heat treatment 334
Drawable charge quantity, starter batteries 1482
DRIE process, micromechanics 1720
Drift current, sensors 1777
Drive by wire, mechatronics 1641
Drive efficiency, energy demand for drives 514
Drive power, alternators 1499
Drive programs, transmission control 632
Drive slip, wheel-slip control systems 1260
Drive structure, hybrid drives 998
Drive torque 800
Drive types, drivetrain with internal-combustion engine 598
Driver identification, interior-monitoring systems 1932
Driver intervention prompt, ACC 1892
Driver support, driver-assistance systems 1806
Driver workload 1304
Driver's cab, vehicle bodies, commercial vehicles 1328
Driver-assistance systems, driver assistance and sensor technology 1804
Driver-drowsiness detection, Information and warning systems 1900

Driver-drowsiness detection, interior-monitoring systems 1932
Drive-shaft gear, differentials 626
Drivetrain of an electric vehicle, control of electric drives 984
Drivetrain operating strategy, control of a hybrid drive 1012
Drivetrain topology, hybrid drives 998
Drivetrain with internal-combustion engine 594
Driving dynamics - commercial vehicles, motor-vehicle dynamics 493
Driving maneuver, with and without ESC 1274
Driving noise, vehicle acoustics 530
Driving safety, chassis systems 1110
Driving-condition estimation, ESC 1276
Driving-dynamic driver steering recommendation 1304
Driving-dynamics control system for commercial vehicles 1287
Driving-dynamics test procedures, vehicle physics 498
Driving-situation detection, transmission control 633
Drop center, wheels 1156
Drop-center rim, wheels 1157
Drop-in fuels, synthetic fuels 583
Droplet formation, gasoline engine 648
Droplet vaporization, gasoline engine 648
Dropping point, lubricants 543
Drum brakes, wheel brakes 1256
Dry boiling point, brake fluids 557
Dry clutch, power take-up elements 601
Dry storage, starter batteries 1483
Dry-sump lubrication, reciprocating-piston engine 688
Dual clutch, power take-up elements 602
Dual spray, intake-manifold fuel injector 847
Dual-battery electrical systems, vehicle electrical systems for commercial vehicles 1468
Dual-circuit version, steering 1204
Dual-clutch transmissions, step-up gear units 618
Dual-headlamp system, lighting equipment 1340
Dual-mass flywheel, power take-up elements 606
Dual-mass flywheels, reciprocating-piston engine 689
Dual-spark ignition coil, ignition coil 863
Dual-stage controlled supercharging, complex supercharging systems 785
Dual-voltage electrical systems, vehicle electrical systems for commercial vehicles 1468
Duct shutoff flap, air-intake system 765
Ductility, material parameters 261
Dummy pedestrian, active pedestrian protection 1921
Dummy-head measurement technique, vehicle acoustics 532
Duplex brake, drum brake 1257
Durability, CARB legislation 1045
Durability, emission-control legislation for heavy commercial vehicles 1064, 1065, 1066
Durability, EPA legislation 1049
Durability, Japanese legislation 1056
Dust preseparation, air-intake system 769
Dwell period, ignition 861
Dwell time, λ control 879
Dwell-period determination, ignition 861
Dynamic cornering headlamps, lighting equipment 1355
Dynamic Coupling Torque at Center, ESC 1283
Dynamic Noise Covering, receivers 1799
Dynamic pressure, aerodynamics 516
Dynamic programming, control of a hybrid drive 1012
Dynamic range, lidar-sensor system 1829
Dynamic route guidance, traffic telematics 1802
Dynamic routing, vehicle navigation 1874
Dynamic seals, seal technology 384
Dynamic segment, FlexRay 1698
Dynamic Stability and Traction Control, driving-dynamics control system 1287
Dynamic Stability Control, driving-dynamics control system 1287
Dynamic Steering, steering 1200
Dynamic supercharging 671
Dynamic Trailing Sequence, FlexRay 1700
Dynamic viscosity 699
Dynamic viscosity, flow mechanics 57
Dynamic viscosity, lubricants 541

Dynamic wheel-load fluctuations, motor-vehicle dynamics 489
Dynamics of lateral motion caused by crosswind, motor-vehicle dynamics 483
Dynamics of lateral motion, driving dynamics - commercial vehicles 493
Dynamics of lateral motion, motor-vehicle dynamics 479
Dynamics of linear motion, driving dynamics - commercial vehicles 493
Dynamics of linear motion, vehicle physics 471
Dynamics of vertical motion, motor-vehicle dynamics 486
Dynamo field (DF), 12 V vehicle electrical system 1456

E

E/E architecture blocks, architecture of electronic systems 1671
E/E architecture, architecture of electronic systems 1667, 1674
Early fusion, sensor-data fusion 1862
Early Pole Crash Detection, occupant-protection systems 1417
Easy entry, electrical seat adjustment 1408
Easy-change oil filter, engine lubrication 751
eAxle, electric final-drive unit 1020
e-axle, electric final-drive unit 1020
EC vehicle categories, overall motor-vehicle system 29
eCall, safety systems in the motor vehicle 1411
Eccentric connecting-rod bearing, reciprocating-piston engine 682
Eccentric crankshaft bearing, reciprocating-piston engine 682
Eccentric shaft, high-pressure pump, common rail 937
Echo-sounding method, ultrasonic sensor system 1813
ECMS, control of a hybrid drive 1014
ECU diagnostics 1106
ECU network, automotive software engineering 1656
Eddy-current brake, continuous-operation brake systems 1248
Eddy-current losses, electrical engineering 124
Eddy-current losses, electrical machines 1579
Eddy-current sensors 1726
EEPROM, electronic control unit (ECU) 1629
EFB battery, starter batteries 1485
Effective braking time 1209
Effective luminous flux, lighting equipment 1346
Effective power, reciprocating-piston engine 705
Effective power, thermal engines 644
Effective value, periodic signal 62
Efficiency chain, thermal engines 645
Efficiency Line, alternators 1505
Efficiency map, electrical machines 1568
Efficiency of a headlamp, lighting equipment 1346
Efficiency optimization, alternators 1501
Efficiency, alternators 1501
Efficiency, electric final-drive unit 1027
Efficiency, fuel cell 1037
Efficiency, reciprocating-piston engine 706
Efficiency, thermal engines 643
Efficiency-of-cycle factor, thermal engines 644
E-fuels, fuels 583
EGR cooler, exhaust-gas cooling 744
EGR cooler, exhaust-gas recirculation 674
Eiffel type, wind tunnels 524
Eigenvalue, modal analysis 64
Eigenvector, modal analysis 64
Elastohydrodynamics 700
Elastomer materials, seals 415
Elastomer mounts, chassis systems 1124
Elastomer swelling, brake fluids 558
Elastomer-metal flat seal, seals 395
Elastomers, names 317
Elastomers, plastics 316
Elastomers, properties 318
Electric actuators, air-intake system 766
Electric conductivity, material parameters 258
Electric continuously variable transmission, hybrid drives 1006
Electric dipoles, chemical bonds 190
Electric drive, electrification of the drive 972
Electric drives with fuel cell, electrification of drive 1030
Electric effects in metallic conductors, electrical engineering 131

Electric field strength,
electrical engineering 117
Electric field, electrical engineering 117
Electric final-drive unit,
electrification of drive 1020
Electric fuel pump, common rail 925
Electric fuel pump, fuel supply,
spark-ignition engine 827
Electric fuel pump,
low-pressure diesel fuel supply 916
Electric lock, locking systems 1440
Electric machines 158
Electric motor, electric fuel pump 827
Electric potential,
electrical engineering 117
Electric power generation,
alternators 1490
Electric power,
electrical engineering 112
Electric power-assisted steering,
steering 1198
Electric range,
energy demand for drives 510
Electric suction valve,
high-pressure pump, common rail 941
Electrical activation,
high-pressure fuel injector 850
Electrical activation,
intake-manifold fuel injector 848
Electrical activation,
piezo high-pressure fuel injector 852
Electrical energy density,
electrical engineering 118
Electrical energy management,
12 V vehicle electrical system 1462
Electrical engineering,
basic principles 102
Electrical isolation,
on-board charger 1605
Electrical machines for motor-vehicle
propulsion, electrics for electric and
hybrid drives 1566
Electrical pitting,
lubricating greases 551
Electrical quantities, legal units 38
Electrical seat adjustment, comfort
and convenience functions in
the passenger compartment 1408
Electrical steel sheet,
magnetic materials 292
Electrical steel strip 292
Electrical Supply System subsystem,
control of electric drives 988
Electrical system overview,
control of electric drives 984
Electrically assisted exhaust-gas
turbocharging 673
Electrically driven compressor,
complex supercharging systems 788
Electrically excited synchronous
machine, electrical machines 1573
Electrical-system concepts,
vehicle electrical systems
for commercial vehicles 1471
Electrical-system parameters,
12 V vehicle electrical system 1460
Electrical-system structures,
12 V vehicle electrical system 1458
Electricity, electrical engineering 103
Electricity-generated regenerative fuels,
fuels 583
Electrification of the drive 970
Electrified accessories,
electrification of drive 975
Electrochemical corrosion,
corrosion processes 340
Electrochemical impedance
spectroscopy, corrosion testing 342
Electrochemical measurement principle,
hydrogen sensors 1774
Electrochemical processes,
corrosion protection 345
Electrochemical reaction,
chemistry 203
Electrochemical series of metals,
chemistry 202
Electrochemical series of metals,
corrosion processes 340
Electrochemical testing procedures,
corrosion testing 342
Electrochemistry, chemistry 202
Electrode gap, gasoline engine 649
Electrode gap, spark plug 869
Electrode wear, spark plug 870
Electrodeposits,
deposits and coatings 348
Electrodes, spark plug 867
Electrodes, starter batteries 1478
Electrodynamic principle,
electrical engineering 121
Electrodynamic retarder, continuous-
operation brake systems 1248
Electrohydraulic actuators,
transmission control 635
Electrohydraulic brake,
passenger-car brake system 1224
Electrohydraulic valve timing 670
Electrolyte density,
starter batteries 1481
Electrolyte, starter batteries 1478

Electromagnetic active chassis system 1134
Electromagnetic compatibility 1524
Electromagnetic fields, electrical engineering 125
Electromagnetic fuel injector, intake-manifold fuel injector 846
Electromagnetic high-pressure fuel injector, mixture formation 848
Electromagnetic induction, alternators 1491
Electromagnetic power density, electrical engineering 129
Electromagnetic radiation, optical technology 74
Electromechanical brake booster, passenger-car brake system 1218
Electromechanical parking brake, automatic brake-system operations 1294
Electromechanical parking brake, passenger-car brake system 1222
Electromechanical power-assisted steering, steering 1198
Electromechanical valve gear 669
Electromobility for two-wheelers, electrification of drive 1040
Electromobility, electrification of the drive 970
Electron orbitals, periodic table 182, 185
Electron, elements 182
Electronegativity, chemical bonds 191
Electronic air-processing unit, brake systems, commercial vehicles 1236
Electronic all-around visibility, driver-assistance systems 1808
Electronic ballast unit, Litronic, lighting equipment 1350
Electronic battery sensor, 12 V vehicle electrical system 1465
Electronic brake system 1206
Electronic braking-force distribution, ESC 1278
Electronic control unit (ECU), automotive electronics 1620
Electronic control, distributor injection pumps with rotary-magnet actuator 949
Electronic control, solenoid-valve-controlled distributor injection pumps 950
Electronic control, unit injector 947
Electronic Diesel Control, management for diesel engines 906
Electronic governor, axial-piston distributor pump 949
Electronic horizon 1895
Electronic load-dependent braking-force regulation, wheel-slip control systems 1264
Electronic Stability Control, driving-dynamics control system 1272
Electronic Stability Program 1272
Electronic Stability Program, driving-dynamics control system 1272
Electronic throttle control (ETC) components, cylinder charge 813
Electronic transmission control, drivetrain 630
Electronic vehicle immobilizer, vehicle security systems 1444
Electronically controlled brake system, brake systems, commercial vehicles 1242
Electronically controlled braking-force distribution, Motorcycle Stability Control 1300
Electronics 138
Electron-volt, legal units 37
Electropneumatic valve timing 670
Electrostatic discharge, electromagnetic compatibility 1527, 1532
Element quality, finite-element method 235
Element spectroscopy, elements 187
Elements, chemistry 182
Elements, finite-element method 235
Elliptical polarization, wave optics 77
Elongation at break, elastomer materials 420
Elongation at fracture, material parameters 261
Embeddability, friction bearings 372
Embedded software 1646
EMC development, electromagnetic compatibility 1529
EMC measuring techniques, electromagnetic compatibility 1529
EMC requirements analysis, electromagnetic compatibility 1528
EMC requirements documents, electromagnetic compatibility 1528
EMC simulation, electromagnetic compatibility 1533
EMC specifications, electromagnetic compatibility 1528
EMC test methods, electromagnetic compatibility 1530
EMC testing, electromagnetic compatibility 1533

EMC validation, electromagnetic compatibility 1529
EMC verification, electromagnetic compatibility 1529
EMC-oriented development, electromagnetic compatibility 1528
Emergency brake assist, emergency-braking systems 1916
Emergency lane-keeping support, lane assistance 1909
Emergency release device, brake systems, commercial vehicles 1240
Emergency-braking systems, driver-assistance systems 1916
Emission characteristic, ultrasonic sensor system 1813
Emission legislation for motorcycles, emission-control legislation 1071
Emission limits, CARB legislation 1043
Emission limits, emission-control legislation for heavy commercial vehicles 1062, 1064, 1066
Emission limits, EPA legislation 1047
Emission limits, EU legislation 1051
Emission limits, Japanese legislation 1055
Emission spectrometry, elements 187
Emission testing, EU legislation 1054
Emission-control legislation for heavy commercial vehicles 1062
Emission-control legislation for passenger cars and light commercial vehicles 1043
Emission-control legislation, emission-control and diagnosis legislation 1042
Emission-free vehicles, CARB legislation 1046
Emissions, diesel engine 661
Emissions, gasoline engine 655
Emissivity, thermodynamics 100
E-modulus, material parameters 260
Emulsion, substances 192
EN metallurgy standards 286
Encoding, CAN 1692
End cover, electric fuel pump 827
End of Frame, CAN 1695
Endothermic reaction, chemical thermodynamics 195
End-point dosing, gasoline 567
Energy bands, chemical bonds 190
Energy demand for drives, vehicle physics 506
Energy management for truck electrical systems 1469
Energy of dissociation, chemical bonds 189
Energy storage, brake systems, commercial vehicles 1236
Energy, legal units 37
Energy, mechanics 47
Energy-assisted brake system 1207
Energy-equivalent continuous sound level, acoustics 70
Energy-level diagram, periodic table 185
Energy-saving system, brake systems, commercial vehicles 1234
Energy-supplying device, brake systems 1206
Engine acoustics 533
Engine air-intake filter, air-intake system 755
Engine cooling 736
Engine drag-torque control 1273
Engine drag-torque control, wheel-slip control systems 1264
Engine fuel-consumption map, energy demand for drives 506
Engine lubrication, internal-combustion engines 750
Engine management for natural-gas vehicles 896
Engine management system, spark-ignition (SI) engine 800
Engine oils, lubricants 544
Engine output, reciprocating-piston engine 705
Engine-oil pressure sensor 1758
Engine-oil temperature sensor 1763
Engines fueled by natural gas 894
Enhanced Seat Belt Reminder, safety systems in the motor vehicle 1410
Enhanced understeering control, ESC 1280
Enhancement type, field-effect transistor 147
Enrichment 839
Enthalpy of activation, chemical thermodynamics 195
Enthalpy of evaporation, material parameters 258
Enthalpy of fusion, material parameters 258
Enthalpy of reaction, chemical thermodynamics 195
Enthalpy, chemical thermodynamics 195
Enthalpy, thermodynamics 88
Entropy, chemical thermodynamics 196

Entropy, thermodynamics 88
Entry angle, belt drives 724
Envelope requirement, tolerances 354
Envelope separator,
starter batteries 1478
Environment detection,
blind-spot assistant 1912
Environment detection,
emergency-braking systems 1916
Environment detection,
lane assistance 1910
Environment detection,
lane-change assistant 1913
Environment perception,
lidar-sensor system 1825
Environment-sensor system,
driver-assistance systems 1808
EOBD, on-board diagnostics 1095
EOL programming,
electronic control unit (ECU) 1639
EP lubricants 538
EPA legislation,
emission-control legislation 1047
EPA OBD, on-board diagnostics 1095
EPA vehicle categories,
overall motor-vehicle system 30
Epichlorohydrin rubber,
elastomer materials 416
Epipoly layer, micromechanics 1721
Epitaxy, micromechanics 1721
Epoxides, plastics 325
Epoxy adhesive,
adhesive technologies 456
Epoxy resins, plastics 319
Equations of state, thermodynamics 89
Equilibrium constant,
acids in water 198
Equilibrium constant,
law of mass action 197
Equipment mechanics, inverters 1591
Equivalence factor,
control of a hybrid drive 1014
Equivalent consumption
minimization strategy,
control of a hybrid drive 1012
Error detection, CAN 1696
Error handling, CAN 1696
Error/fault detection, diagnostics 1091
Error/fault handling, diagnostics 1091
Error/fault storage, diagnostics 1091
ESC-based regenerative braking system,
electrification of drive 977
ESI[tronic] 1552
Ethanol, alcohol operation 901
Ethernet protocol 1704
Ethernet, buses in motor vehicles 1703
Ethers, fuels 584
Ethylene acrylate rubber,
elastomer materials 416
Ethylene propylene diene rubber,
elastomer materials 416
ETN nomenclature,
starter batteries 1482
EU legislation, emission-control
legislation for heavy commercial
vehicles 1064
EU legislation, emission-control
legislation 1045, 1050
EU tire label 1185
EU/ECE legislation, Emission
legislation for motorcycles 1073
Euler's rope-friction formula,
mechanics 54
Euler's formula, mathematics 211
Euro 5,
emission-control legislation 1051
Euro 6d-temp,
emission-control legislation 1052
Euro NCAP assessment,
emergency-braking systems 1918
Euro NCAP, safety systems
in the motor vehicle 1415
Euronozzle, LPG operation 892
European Steady-State Cycle,
test cycles for heavy commercial
vehicles 1069
European test cycle,
emission-control legislation 1059
European Transient Cycle, test cycles
for heavy commercial vehicles 1069
European Type Number,
starter batteries 1482
Evaluation of the traffic situation,
emergency-braking systems 1917
Evaporative emissions, emission
legislation for motorcycles 1072, 1074
Evaporative emissions,
evaporative-emissions test 1087
Evaporative losses,
evaporative-emissions test 1087
Evaporative-emissions control system,
fuel supply 824
Evaporative-emissions test, exhaust-gas
measuring techniques 1087
Evaporator, climate control 1395
Evaporator, LPG operation 892
Event Data Recorder, safety systems
in the motor vehicle 1411
Excess-air factor, mixture formation,
spark-ignition engines 838

Excessive noise, acoustics 70
Excitation circuit, alternators 1495
Excitation current,
12 V vehicle electrical system 1455
Excitation diodes, alternators 1494
Excitation field, alternators 1491
Excitation winding, alternators 1492
Excursion force,
snap-on connections 448
Exergy, thermodynamics 92
Exhaust brake, continuous-operation brake systems 1246
Exhaust camshaft,
variable valve timing 667
Exhaust closes, charge cycle 665
Exhaust emissions,
engines fueled by natural gas 897
Exhaust flap, continuous-operation brake systems 1246
Exhaust flow meter, exhaust-gas measuring techniques 1084
Exhaust manifold,
exhaust-gas system 792
Exhaust opens, charge cycle 665
Exhaust passage, charge cycle 665
Exhaust ports, 2-stroke cycle 666
Exhaust stroke, charge cycle 664
Exhaust System, Motronic 803
Exhaust valve, charge cycle 664
Exhaust valve,
reciprocating-piston engine 685
Exhaust-brake systems, continuous-operation brake systems 1246
Exhaust-camshaft adjustment,
catalytic-converter heating 877
Exhaust-camshaft adjustment,
variable valve timing 667
Exhaust-gas categories,
CARB legislation 1044
Exhaust-gas categories,
EPA legislation 1048
Exhaust-gas cooling,
engine cooling 744
Exhaust-gas emissions,
emission legislation
for motorcycles 1071
Exhaust-gas emissions,
internal-combustion engines 716
Exhaust-gas flaps,
exhaust-gas system 797
Exhaust-gas mass flow,
exhaust-gas turbocharging 672
Exhaust-gas measuring devices,
exhaust-gas measuring
techniques 1081
Exhaust-gas measuring techniques,
emission-control and diagnosis
legislation 1078
Exhaust-gas noise 791
Exhaust-gas recirculation system
diagnosis, on-board diagnostics 1100
Exhaust-gas recirculation 674
Exhaust-gas recirculation,
cylinder charge 814
Exhaust-gas recirculation,
Electronic Diesel Control 914
Exhaust-gas recirculation,
exhaust-gas cooling 744
Exhaust-gas recirculation,
gasoline engine 651
Exhaust-gas sample bag, exhaust-gas measuring techniques 1080
Exhaust-gas system 790
Exhaust-gas temperature sensor 1763
Exhaust-gas testing, exhaust-gas measuring techniques 1078
Exhaust-gas treatment,
diesel engines 962
Exhaust-gas treatment,
exhaust-gas system 790
Exhaust-gas treatment, management
for spark-ignition engines 874
Exhaust-gas turbocharging 672
Exothermic reaction,
chemical thermodynamics 195
Expanded metal grid,
starter batteries 1479
Expansion fit, frictional joints 433
Expansion reservoir,
passenger-car brake system 1221
Expansion tank, water cooling 739
Expansion valve, climate control 1395
Expansion-element-regulated
thermostat, water cooling 741
Expectation value,
technical statistics 230
Experimental modal analysis 65
Explosive rivet, riveting 458
Exponential equation,
NTC thermistor 1761
Exponential function, mathematics 206
Exposure frequency, tire grip 1178
Exposure tests, corrosion testing 343
Expressway beam,
lighting equipment 1357
Extended current output,
starter batteries 1486
Extended hump, wheels 1156
Extension springs, springs 368
Extensive state variables 87

External dimension,
vehicle bodies, passenger cars 1318
External excitation,
12 V vehicle electrical system 1455
External excitation, alternators 1491
External exhaust-gas recirculation,
cylinder charge 814
External mixture formation,
gasoline engine 646
External mixture formation,
spark-ignition engines 841
External running resistances,
energy demand for drives 507
External seals, pneumatic seals 406
Externally chargeable hybrids,
hybrid drives 998
Externally rechargeable hybrid,
electrification of the drive 971
Extraneously caused dirtying 523
Extrema, mathematics 216
Extrinsic photoelectric effect,
sensors 1776
Eye safety, lidar-sensor system 1829

F

Facet-type reflector,
lighting equipment 1347
Fading, receivers 1799
Failure probability,
technical statistics 229
Failure rate, technical statistics 229
Fallback level, brake systems
for passenger cars 1228
Fan activation,
Electronic Diesel Control 914
Fan control, cooling-air fan 740
Farad, SI units 34
Far-infrared systems,
night vision systems 1876
Fast Ethernet 1703
Fast FMCW modulation,
radar-sensor system 1820
Fast light-off, λ sensors 1768
Fastener forces,
threaded fasteners 439
Fastener torques,
threaded fasteners 439
Fatigue strength under reversed bending
stress, material parameters 261
Fatigue strength, friction bearings 372
Fatigue-limit diagrams, springs 367
Fatty acid methyl esters,
diesel fuels 575
Fault code, diagnostics 1091
Fault type, diagnostics 1091
Fault-memory entries,
ECU diagnostics 1106
Fault-tolerant communication
network, architecture of electronic
systems 1671
Fault-tolerant vehicle electrical
system, architecture of electronic
systems 1672
Feather-key couplings,
positive or form-closed joints 426
Feature model, architecture
of electronic systems 1675
Feature vector,
sensor-data fusion 1860
Federal Smoke Cycle, test cycles
for heavy commercial vehicles 1068
Felt, air-intake system 757
FEM examples,
finite-element method 236
FEM program system,
finite-element method 234
Fermat's principle, geometric optics 75
Ferrimagnets, magnetic materials 290
Ferrite, steels 271
Ferritic structure, steels 271
Ferroelectric Random Access Memory,
electronic control unit (ECU) 1628
Ferromagnetic materials,
electrical engineering 122
Ferromagnets, magnetic materials 290
Ferrous materials,
metallic materials 271
Fiber composite materials 303
Fiber-optic technology,
lighting equipment 1366
Field characteristic impedance,
electrical engineering 129
Field-effect transistors, electronics 146
Filament lamp, lighting equipment 1336
Fill simulation,
two-component seals 408
Fillers, plastics 322
Fillers, radial tire 1171
Filling duct, air-intake system 765
Filling-station network,
engines fueled by natural gas 894
Fill-volume restrictor,
LPG operation 887
Filter element, climate control 1393
Filter elements, air-intake system 756
Filter fineness, gasoline filter 826
Filter fleece, climate control 1393
Filter media, cabin filters 1397
Filter media, engine lubrication 752

Filter medium, gasoline filter 826
Filter method,
diesel smoke-emission test 1085
Filter smoke number,
diesel smoke-emission test 1087
Filterability, diesel fuel 572
Filtration of particulates, exhaust-gas
treatment, diesel engines 963
Final compression pressure,
reciprocating-piston engine 715
Final compression temperature,
reciprocating-piston engine 715
Final-drive differential, differentials 626
Finally annealed state,
electrical steel strip 292
Finite-element analysis 234
Finite-element method 234
Fire point 538
Fire-polishing, automotive glazing 1374
Firing pellet,
occupant-protection systems 1421
Firing sequence,
reciprocating-piston engine 691
Firing voltage, ignition 854
First law, thermodynamics 88
First-order reaction,
reaction kinetics 197
Fischer-Tropsch process,
diesel fuel 577
Five-link independent wheel suspension,
wheel suspensions 1153
Fixed-caliper brake, disk brakes 1251
Flame front, ignition 854
Flame glow plug,
start-assist systems 959
Flame ionization detector, exhaust-gas
measuring techniques 1083
Flame propagation,
gasoline engine 650
Flame soldering, soldering 455
Flame start system with burner chamber,
start-assist systems 958
Flame start systems,
start-assist systems 958
Flame velocity, gasoline engine 651
Flange height, wheels 1156
Flange-to-flange width, wheels 1156
Flank diameter, threaded fasteners 437
Flaps, climate control 1392
Flash lidar, lidar-sensor system 1835
Flash memory,
electronic control unit (ECU) 1629
Flash point 538
Flash point, diesel fuel 572
Flash-butt welding, welding 451
Flashing frequency,
lighting equipment 1359
Flat glass, automotive glazing 1374
Flat hump, wheels 1156
Flat round-head rivet, riveting 458
Flat seal seat, spark plug 867
Flat seals, seals 395
Flat-base rim, wheels 1158
Flatbed van, vehicle bodies,
commercial vehicles 1324
Flat-seat pressure-control valve,
transmission control 637
Flat-tube oil coolers, oil cooling 746
Flat-tube/corrugated-fin systems,
oil cooling 745
Flat-tube/corrugated-fin systems,
water cooling 737
Fleet average, CARB legislation 1045
Fleet average, EPA legislation 1049
Fleet consumption,
CARB legislation 1046
Fleet consumption, EPA legislation 1049
Fleet consumption,
Japanese legislation 1056
Flex spline, steering 1201
Flex-fuel components,
alcohol operation 904
Flex-fuel concept,
alcohol operation 903
Flex-fuel operation, alternative
gasoline-engine operation 901
Flex-fuel, alcohol operation 902
Flexibility at low temperature,
elastomer materials 422
Flexible cable window lift,
power-window systems 1404
Flexible input module,
electronic control unit (ECU) 1633
Flexible programmable output module,
electronic control unit (ECU) 1634
Flexing work, rolling resistance,
tires 1180
Flex-pole actuator,
locking systems 1440
FlexRay Consortium 1648
FlexRay, buses in motor vehicles 1697
Float angle, chassis systems 1120
Float angle, vehicle motion 462
Float-glass process,
automotive glazing 1374
Floating-caliper brake, disk brake 1251
Floating-car principle,
traffic telematics 1802
Floating-phone principle,
traffic telematics 1802

Flow direction, flow meters 1747
Flow improvers, diesel fuel 574
Flow mechanics, fluid mechanics 56, 57
Flow meters, sensors 1746
Flow pattern, aerodynamics 517
Flow pressure, lubricants 538
Flow profile, gasoline engine 652
Flow velocity, flow mechanics 57
Flow, laminar 57
Flow, turbulent 57
Flow-forming process 1164
Flow-optimized geometry, natural-gas injector 898
Flows, gasoline engine 653
Flow-type pump, electric fuel pump 828
Fluid, flow mechanics 57
Fluid-friction clutch, water cooling 740
Fluorescent lamps, lighting equipment 1336
Fluorescent radiation, elements 187
Fluorine rubber, elastomer materials 417
Fluorosilicone rubber, elastomer materials 417
Flux density, electrical engineering 119
Fluxes, soldering 454
Flywheel masses, reciprocating-piston engine 689
FMCW measurement method, lidar-sensor system 1834
FMCW modulation, radar-sensor system 1818
FMCW radar, radar-sensor system 1818
Focal point, geometric optics 76
Focusing optics, laser welding 453
Fog lamps, lighting equipment 1353
Folded/pleated filter medium, air-intake system 756
Footprint, chassis systems 1120
Footprint, tires 1176
Force application, threaded fasteners 439
Force blending, electrification of drive 978
Force impulse, mechanics 49
Force measurement, sensors 1765
Force sensor 1765
Force transmission, tires 1176
Force, legal units 37
Force, mechanics 46
Force-application factor, threaded fasteners 439
Forced oscillations 61, 63
Forced-feed lubrication system, engine lubrication 750
Forced-feed lubrication, reciprocating-piston engine 688
Force-limitation device, power-window systems 1404
Forged rail, common rail 943
Forged wheels, wheels 1161
Fork shape, rpm sensors 1739
Forked-tube coolers, oil cooling 746
Form deviations, tolerances 354
Form drag, aerodynamics 517
Form tolerance, tolerances 356
Formation of stress cracks, plastics 310
Forms of energy, mechanics 49
Forms of energy, thermodynamics 87
Formulation function, finite-element method 235
Forward collision warning, advanced rider-assistance systems for two-wheelers 1931
Forward converter, DC-DC converters 1595
Four-circuit protection valve, brake systems, commercial vehicles 1235
Fourier series, vibrations and oscillations 61
Fourier thermal-conduction equation, thermodynamics 97
Fourier transform, mathematics 220
Four-stroke cycle, charge cycle 664
Four-way catalytic converter, catalytic exhaust-gas treatment, gasoline engine 881
Four-zone climate control 1393
Fracture contraction, material parameters 261
Fracture separation, reciprocating-piston engine 679
Fracture toughness, material parameters 262
Frame bitstream, FlexRay 1700
Frame check, CAN 1696
Frame End Sequence, FlexRay 1700
Frame ID, FlexRay 1699
Frame Start Sequence, FlexRay 1700
Frame, alternators 1492
Frames, vehicle communication networks 1682
Free electrons, electrical engineering 103
Free forces, reciprocating-piston engine 697
Free oscillations 63
Free run length, chain drives 733
Free-cutting steel, metallic materials 272

Free-form reflector,
lighting equipment 1346, 1348, 1364
Free-wheeling diode, alternators 1496
Free-wheeling diode, inverters 1583
Free-wheeling diodes, electronics 142
Freeze frame, diagnostics 1092
Freezing threshold,
starter batteries 1482
Frequency generation,
radar-sensor system 1819
Frequency modulation,
radio and TV broadcasting 1795
Frequency ranges,
rolling resistance, tires 1181
Frequency 67
Frequency, technical statistics 226
Frequency,
vibrations and oscillations 60
Fresh A/F mixture, cylinder charge 812
Fresh air, climate control 1392
Fresh charge, charge cycle 664
Friction bearings, machine parts 370
Friction belt drives,
internal-combustion engines 720
Friction losses,
electrical machines 1579
Friction losses, radial shaft seal 405
Friction modifiers, lubricants 539
Friction on the wedge, mechanics 54
Friction steering system, steering 1193
Friction, mechanics 53
Friction,
reciprocating-piston engine 699
Frictional force, mechanics 53
Frictional joints,
detachable connections 430
Friend key, locking systems 1442
Frits, starter batteries 1478
Front airbags,
occupant-protection systems 1419
Front electrode, spark plug 867
Front wiper drive, windshield,
rear-window and headlamp
cleaning systems 1381
Front wiper systems, windshield,
rear-window and headlamp
cleaning systems 1380
Frontal-area measuring system,
wind tunnels 526
Front-end camera, parking and
maneuvering systems 1886
Front-passenger airbag deactivation,
occupant-protection systems 1423
Front-wheel drive, drivetrain with
internal-combustion engine 598
FTP 72 test cycle, test cycles 1059
FTP 75 test cycle, test cycles 1057
Fuel cell, electrification of drive 1030
Fuel consumption, emission-control
legislation for heavy commercial
vehicles 1065
Fuel consumption,
energy demand for drives 506
Fuel consumption,
reciprocating-piston engine 703
Fuel consumption, tire label 1186
Fuel cooler,
low-pressure diesel fuel supply 920
Fuel delivery, LPG operation 887
Fuel droplets, gasoline engine 648
Fuel filter, diesel 922
Fuel filter, fuel supply 825
Fuel filtering, common rail 926
Fuel grades, gasoline 563
Fuel injection, common rail 927
Fuel injectors, Mixture formation,
spark-ignition engines 846
Fuel metering,
axial-piston distributor pump 950
Fuel rail, fuel supply,
gasoline engine 833
Fuel standards, diesel fuel 569
Fuel standards, gasoline 563
Fuel supply and delivery,
gasoline direct injection 822
Fuel supply and delivery,
intake-manifold injection 820
Fuel supply and delivery,
low-pressure diesel fuel supply 916
Fuel supply, common rail 925
Fuel supply,
low-pressure diesel fuel supply 916
Fuel supply, management
for gasoline engines 820
Fuel System, Motronic 803
Fuel vaporization, gasoline engine 647
Fuel volatility, gasoline 565
Fuel-balancing control, common rail 928
Fuel-balancing control,
Electronic Diesel Control 913
Fuel-cell electric drive, fuel cell 1030
Fuel-cell stack, fuel cell 1034
Fuel-cell system, fuel cell 1034
Fuel-consumption requirement,
emission-control legislation for heavy
commercial vehicles 1064, 1066
Fuel-conversion factor,
thermal engines 645
Fuel-delivery control on the suction side,
common rail 927

Fuel-delivery control,
high-pressure pump, common rail 939
Fuel-injection system,
distributor injection pump 953
Fuel-level sensor 1726
Fuel-pressure attenuator, fuel supply,
spark-ignition engine 836
Fuel-pressure attenuator,
gasoline direct injection 832
Fuel-pressure regulator,
fuel supply 820
Fuel-pressure regulator,
fuel supply, spark-ignition engine 835
Fuel-pressure sensor 1758
Fuel-quantity mean-value adaptation,
common rail 928
Fuels, operating fluids 560
Fuel-supply control valve,
gasoline direct injection 832
Fuel-supply module, fuel supply,
spark-ignition engine 829
Fuel-supply pump,
axial-piston distributor pump 948
Fuel-temperature sensor 1763
Full CAN controllers 1696
Full drives, belt drives 723
Full hybrid,
electrification of the drive 971
Full hybrid, hybrid drives 996
Full lift,
piezo high-pressure fuel injector 852
Full speed range,
Adaptive Cruise Control 1894
Full-engine operation,
cylinder shutoff 818
Full-flow oil filters,
engine lubrication 752
Fullpass applications, automotive
software engineering 1658
Full-wave rectification,
alternators 1493
Fully developed deceleration,
brakes 1210
Fully loaded systems,
suspension systems 1133
Fully synthetic filter media,
air-intake system 757
Fully variable valve gear 669
Fully-finished, electrical steel strip 292
Function block, MOST 1710
Function classes, MOST 1710
Function network, automotive
software engineering 1656
Functional design glazing,
automotive glazing 1376
Functional model, architecture
of electronic systems 1675
Functional safety,
electronic control unit (ECU) 1639
Functional safety, inverters 1591
Functional system overview,
control of electric drives 985
Functional units of a motor vehicle,
overall motor-vehicle system 28
Functionalities of hybrid drives,
hybrid drives 992
Functions, mathematics 204
Furrow depth,
reciprocating-piston engine 701
Fusion curve, water 193
Fusion types,
sensor-data fusion 1862
Fusion welding, welding 452
Future of automated driving,
driver-assistance systems
and automated driving 1934

G

Galvanomagnetic effects,
electrical engineering 133
Gamma angle,
intake-manifold fuel injector 848
Gap dimension, molded seal 394
Gas carburizing, heat treatment 337
Gas constant, thermodynamics 89
Gas exchange noise,
vehicle acoustics 533
Gas exchange, charge cycle 664
Gas inflators,
occupant-protection systems 1421
Gas PEMS, exhaust-gas
measuring techniques 1084
Gas rail,
engines fueled by natural gas 896
Gas sensors 1766
Gas shutoff valve, natural-gas
pressure-control module 899
Gas spring with bellows,
suspension 1130
Gas springs, suspension 1128
Gas-discharge lamps,
lighting equipment 1336, 1349
Gaseous fuels, properties 587
Gaseous LPG injection,
LPG operation 889
Gas-exchange TDC, charge cycle 664
Gas-force components,
reciprocating-piston engine 691
Gasoline direct injection 843

Gasoline direct injection, gasoline engine 648
Gasoline engine, mixture formation 646
Gasoline filter, fuel supply 825
Gasoline, fuels 563
Gas-shielded metal-arc welding, welding 452
Gas-to-Liquid, fuels 577
Gateway, architecture of electronic systems 1667
Gaussian bell-shaped curve, technical statistics 230
Gaussian effect, position sensors 1729
Gaussian normal distribution, technical statistics 229
GC FID, exhaust-gas measuring techniques 1083
Gear pump, common rail 925
Gear pump, low-pressure diesel fuel supply 917
Gear ratio, drivetrain with internal-combustion engine 597
Gears, step-up gear units 608
Gear-shift elements, step-up gear units 611
Gear-shift rasping, step-up gear units 612
Gearshift sequence, transmission control 632
Gearshift sleeve, step-up gear units 612
Gel battery, starter batteries 1485
Gel-type greases, lubricants 539
General Timer Module, electronic control unit (ECU) 1626
General tolerances, tolerances 356
General-purpose battery, 12 V vehicle electrical system 1459
Generation of grip, tires 1177
Generator mode, hybrid drives 993
Geodetic pressure, hydrostatics 56
Geometric mean, technical statistics 227
Geometric optics 74
Geometric range of a headlamp, lighting equipment 1345
Geometrical deviations, tolerances 354
Geometrical Product Specifications, tolerances 354
Giant magnetoresistive effect, electrical engineering 136
Giant magnetoresistive effect, position sensors 1737
GIDAS, driver-assistance systems 1805
Gigabit Ethernet 1703
Glare-free high beam, intelligent headlamp control 1929
Glare-free high beam, lighting equipment 1357
Glass behavior, tire grip 1178
Glass formation temperature, glass 303
Glass transition temperature, plastics 316
Glass, automotive glazing 1374
Glass, materials 303
Glass-fiber fleece, starter batteries 1484
Glazing, automotive glazing 1375
Global Positioning System 1872
Global warming potential, refrigerants 591
Glow control unit, preheating systems 957
Glow plugs, preheating systems 954
Glycol-ether fluids, brake fluids 558
GMR effect, electrical engineering 136
GMR sensor, rpm sensors 1739
GMR sensors 1737
GMR sensors, electrical engineering 136
Gold deposits, deposits and coatings 349
Göttingen type, wind tunnels 524
Graded-index fiber, optical fibers/waveguides 83
Gradient control, Motorcycle Stability Control 1299
Gradient sensors, rpm sensors 1741
Gradient structure, air-intake system 757
Gradient, running resistances 468
Gradual braking 1208
Grain oriented electrical steel strip 292
Graph diagram, energy demand for drives 506
Graphics modules, display and control 1787
Graphite, cast iron materials 277
Graphite, lubricants 539
Grashof number, thermodynamics 99
Gravimetric process, exhaust-gas measuring techniques 1083
Gravitation vector, Motorcycle Stability Control 1297
Gravity brake system 1207
Gray cast iron with lamellar graphite, metallic materials 277
Gray cast iron with spheroid graphite, metallic materials 277

Gray cast iron with vermicular graphite, metallic materials 277
Gray cast iron, metallic materials 277
Gray, SI units 34
Grid approaches, sensor-data fusion 1861
Grid heater, start-assist systems 961
Grid structure, starter batteries 1479
Grinding noises, vehicle acoustics 535
Groove filling, O-ring 388
Grooved sliding bearings, friction bearings 373
Gross calorific value, fuels 562
Ground electrode, spark plug 867
Groundhook controllers, shock absorbers 1143
Groups, periodic table 182
Guide band, pneumatic seals 406
Guide bands, rod and piston sealing systems 413
Guide pin, positive or form-closed joints 429
Guide rail, chain drives 735
Guide rings, rod and piston sealing systems 412
Guides, rod and piston sealing systems 413
Gusset plate, vehicle bodies, commercial vehicles 1328

H

Half-bridge, inverters 1583
Half-engine operation, cylinder shutoff 817
Half-life, elements 187
Half-value width, vibrations and oscillations 61
Hall constant, electrical engineering 133
Hall effect, electrical engineering 133
Hall effect, position sensors 1729
Hall sensor, rpm sensors 1738
Hall sensors, position sensors 1730
Hall voltage, position sensors 1729
Halogen double-filament bulb, lighting equipment 1340
Halogen filling, lighting equipment 1336
Halogen lamps, lighting equipment 1336
Handbrake, passenger-car brake system 1216
Handling characteristics, driving dynamics - commercial vehicles 495
Haptic channel, display and control 1785
Hard chrome, deposits and coatings 349
Hard soldering, soldering 454
Hardcase cell, batteries for electric and hybrid drives 1611
Hardener, adhesive technologies 455
Hardening depth, heat treatment 334
Hardening, adhesive technologies 455
Hardening, heat treatment 331
Hard-material coatings, deposits and coatings 352
Hardness testing, heat treatment 328
Hardness tests, elastomer materials 419
Hardness, elastomer materials 419
Hardness, heat treatment 328
Hardness, material parameters 262
Hardness-testing methods, heat treatment 328
Hardware accelerator, electronic control unit (ECU) 1626
Hardware encapsulation, automotive software engineering 1649
Hardware model, architecture of electronic systems 1677
Harmonic drive, steering 1201
Harmonic factor, vibrations and oscillations 62
Harmonic, inverters 1583
Hazard recognition time, motor-vehicle dynamics 475
Hazard-warning and turn-signal flashers, lighting equipment 1360
HC adsorption, air-intake system 758
HCCI combustion process, diesel engine 662
HCI system, exhaust-gas treatment, diesel engines 964
Header CRC, FlexRay 1699
Header, FlexRay 1699
Header, LIN 1701
Headlamp adjustment, lighting equipment 1367, 1368
Headlamp aiming device, lighting equipment 1369
Headlamp aiming devices, lighting equipment 1371
Headlamp cleaning systems, windshield, rear-window and headlamp cleaning systems 1390
Headlamp leveling control, lighting equipment 1367, 1368
Headlamp range 1345
Headlamp systems, lighting equipment 1340

Headlamp technology, lighting equipment 1346
Headlamp, lighting equipment 1346
Headlamps for visual aim, lighting equipment 1345
Headlamps with facet-type reflectors, lighting equipment 1347
Head-on collision, integral safety 1426
Head-up display, display and control 1791
Healing, diagnostics 1091
Hearing dynamics, acoustics 67
Heat capacity, material parameters 258
Heat dissipation, electrical machines 1580
Heat dissipation, inverters 1592
Heat losses due to current, electrical machines 1579
Heat of evaporation, chemical thermodynamics 196
Heat of evaporation, material parameters 258
Heat of fusion, material parameters 258
Heat of hydration, chemical bonds 189
Heat of reaction, chemical thermodynamics 195
Heat of solvation, chemical bonds 189
Heat range, spark plug 867
Heat resistance, elastomer materials 415
Heat transfer, thermodynamics 96
Heat transmission, thermodynamics 98
Heat treatment of metallic materials 328
Heat treatment, processes 331
Heat, legal units 38
Heat, thermodynamics 87
Heat-dependence behavior, tire grip 1178
Heated laminated safety glass, automotive glazing 1377
Heating circuit, climate control 1395
Heating filament, preheating systems 956
Heating press, tires 1172
Heating, catalytic converter 877
Heat-range code number, spark plug 867
Heat-range reserve, spark plug 868
Heat-transfer coefficient, thermodynamics 99
Heavy-duty batteries, starter batteries 1486
Heel point, vehicle bodies, passenger cars 1318
Helical springs, springs 366
Helical springs, suspension 1127
Helical toothing, step-up gear units 609
Helical travel, starters for passenger cars 1510
Helical-gear drive, steering 1200
Helmholtz resonator, exhaust-gas system 797
Henry, SI units 34
Hertz, SI units 34
Hertzian stress 701
Hertzian stress, mechanics 51
Heterogeneous combustion processes, gasoline engine 648
Heterogeneous mixture formation, gasoline engine 647
Heterogeneous substances, chemistry 192
Heuristics, sensor-data fusion 1860
Hexadecimal system, mathematics 204
Hexahedrons, finite-element method 236
HH split, antilock braking system 1267
HI split, antilock braking system 1267
High beam, Europe, lighting equipment 1340, 1343
High beam, USA, lighting equipment 1344
High Temperature High Shear, lubricants 542
High-beam distribution, lighting equipment 1341
High-capacity road train, vehicle bodies, commercial vehicles 1324
High-capacity van, vehicle bodies, passenger cars 1315
High-cut, receivers 1800
High-end ECUs, motor sport 884
High-energy coils, ignition coil 864
High-frequency interference coupling, electromagnetic compatibility 1531
High-frequency vibrations and oscillations, electromagnetic compatibility 1526
High-level signal interference, receivers 1799
High-lubricity oils 540
Highly automated, remote-controlled parking, parking and maneuvering systems 1888
High-pin-count plug-in connections 1521
High-pressure circuit, high-pressure pump, common rail 941
High-pressure EGR, exhaust-gas recirculation 674
High-pressure fuel injector, mixture formation 848

High-pressure lubricants 539
High-pressure port, high-pressure pump, common rail 938
High-pressure pump, distributor injection pump 948
High-pressure pump, gasoline direct injection 830
High-pressure pump, radial-piston distributor pump 951
High-pressure pumps, common rail 937
High-pressure sensors with metal diaphragm 1759
High-pressure shutoff valve, engines fueled by natural gas 896
High-pressure shutoff valves, engines fueled by natural gas 896
High-pressure solenoid valve, axial-piston distributor pump 950
High-pressure solenoid valve, radial-piston distributor pump 952
High-pressure solenoid valve, unit injector 944
High-pressure switching valve, ESC 1285
High-pressure system, fuel supply 822
High-side switches, inverters 1583
High-speed CAN 1693
High-speed tool steels, metallic materials 272
High-voltage terminal, spark plug 866
Highway assistant, future of automated driving 1936
Highway beam, lighting equipment 1357
Highway pilot, future of automated driving 1937
Hill hold control, automatic brake-system operations 1294
Hold-in winding, starters for passenger cars 1508
Holding-current phase, solenoid-valve injector 933
Hollow rivet, riveting 458
Home-zone park assist, parking and maneuvering systems 1888
Homogeneous mixture formation, gasoline engine 646
Homogeneous operation, gasoline direct injection 844
Homogeneous split, catalytic-converter heating 877
Homogeneous substances, chemistry 192
Homogeneous-charge compression ignition in diesel operation, mixture formation 662
Homogeneous-lean mode, gasoline direct injection 845
Homogeneous-split mode, gasoline direct injection 845
Homogeneous-stratified mode, gasoline direct injection 845
Homologation, radar-sensor system 1822
Honing pattern 701
Honing, reciprocating-piston engine 701
Hooke's law, mechanics 51
Hooke's straight line, mechanics 51
Host, vehicle communication networks 1684
Hot aging, heat treatment 336
Hot soak, evaporative-emissions test 1088
Hot spark plug 868
Hot-film air-mass meters, flow meters 1748
Hot-forged rail, common rail 943
Hot-immersion deposits, deposits and coatings 350
Hot-soak losses, evaporative-emissions test 1088
Hot-water unit, wind tunnels 527
Hot-wire air-mass meters, flow meters 1747
Hot-work steels, metallic materials 272
Housing bore, radial shaft seal 402
Housing oil filter, engine lubrication 751
HTHS viscosity, lubricants 542
Huang's algorithm, shock absorbers 1144
Hub, Ethernet 1704
Hub, vehicle communication networks 1689
Hubload, belt drives 721
HUD glazing, automotive glazing 1378
Human eye 79
Hump, wheels 1156
Huygens' principle, wave optics 78
HV battery, vehicle electrical systems for hybrid and electric vehicles 1561
HV vehicle electrical system, vehicle electrical systems for hybrid and electric vehicles 1561
Hybrid control units, electronic control unit (ECU) 1636
Hybrid control, hybrid drives 1010
Hybrid cycle, test cycles 1059
Hybrid drive, electrification of the drive 973
Hybrid drives, electrification of the drive 990

Hybrid driving, hybrid drives 992
Hybrid filter, climate control 1393
Hybrid functionalities, hybrid drives 992
Hybrid radiator protection agents, coolants 555
Hybrid supercapacitors, electromobility 1618
Hybrid topology, vehicle communication networks 1690
Hybrid vehicle, hybrid drives 990
Hydraulic brake assistant, automatic brake-system operations 1292
Hydraulic brake boosters, passenger-car brake system 1217
Hydraulic control, transmission control 634
Hydraulic cylinders, chassis systems 1133
Hydraulic damping, chassis systems 1130
Hydraulic fallback level, brake systems for passenger cars 1226
Hydraulic fallback level, electrohydraulic brake 1225
Hydraulic hollow-jacket spring collet, frictional joints 434
Hydraulic modulator, antilock braking system 1265
Hydraulic modulator, electrohydraulic brake 1225
Hydraulic modulator, passenger-car brake system 1221
Hydraulic module, brake systems for passenger cars 1227
Hydraulic piston seals, seals 413
Hydraulic power-assisted steering, steering 1196
Hydraulic retarder, continuous-operation brake systems 1247
Hydraulic rod seals, seals 411
Hydraulic schematic, Motorcycle Stability Control 1301
Hydraulic telescopic shock absorber, shock absorbers 1136
Hydraulic unit, antilock braking system 1265
Hydraulic unit, ESC 1285
Hydraulic unit, traction control system 1266
Hydraulic-pressure sensor 1758
Hydrocarbon Injection, exhaust-gas treatment, diesel engines 964
Hydrocarbons, exhaust-gas emissions 717
Hydro-crack oils, lubricants 539
Hydrodynamic friction bearings 371
Hydrodynamic retarder, continuous-operation brake systems 1247
Hydrodynamic torque converter, power take-up elements 603
Hydrodynamics, lubricants 543
Hydrogen bonds, chemical bonds 191
Hydrogen supply, fuel cell 1034
Hydrogen, fuels 582
Hydrogenated nitrile butadiene rubber, elastomer materials 416
Hydrographs, traffic telematics 1802
Hydropneumatic spring, suspension 1129
Hydropneumatic springs, suspension 1126, 1130
Hydropneumatic suspension system, suspension 1134
Hydropulsers, vehicle acoustics 532
Hydrostatic friction bearings 370
Hydrostatic press, hydrostatics 56
Hydrostatic pressure, hydrostatics 56
Hydrostatics, fluid mechanics 56
Hydroxide ions, acids in water 199
Hydroxonium ions, acids in water 198
Hypergeometric distribution, technical statistics 228
Hypoid drive, differentials 626
Hysteresis loop, electrical engineering 122
Hysteresis loss, electrical engineering 122
Hysteresis losses, electrical machines 1579

I

iBolt, force sensor 1765
iBolt, occupant-protection systems 1422
iBooster, passenger-car brake system 1218
IC test methods, electromagnetic compatibility 1530
Ice flaking point, coolants 554
Ideal gas, thermodynamics 89
Ideal nozzle, natural-gas injector 898
Ident Field, LIN 1701
Identification lamps, lighting equipment 1363
Identifier field, LIN 1701
Identifier, CAN 1694
Identifier, circuit diagrams 1549
Identifier, LIN 1702

Idle speed increase,
catalytic-converter heating 877
Idle, FlexRay 1697
Idle_LP, FlexRay 1697
Idler pulleys, Belt drives 724
Idle-speed control,
Electronic Diesel Control 912
Idling, synchronous machine 171
IGBT, electronics 146
Ignition advance step,
knock control 858
Ignition advance waiting time,
knock control 858
Ignition coil, ignition 862
Ignition energy, gasoline engine 650
Ignition energy, ignition 854
Ignition intervals,
reciprocating-piston engine 697
Ignition lag, diesel engine 658
Ignition maps, ignition 856
Ignition point, charge cycle 665
Ignition release, ignition 861
Ignition retard step, knock control 858
Ignition spark, ignition 854
Ignition sparks, gasoline engine 649
Ignition System, Motronic 803
Ignition systems, ignition 860
Ignition TDC, charge cycle 664
Ignition temperature, fuels 562
Ignition voltage, ignition 854
Ignition voltage, spark plug 869
Ignition, gasoline engine 649, 650
Ignition, ignition 854
Ignition-angle adjustment,
catalytic-converter heating 877
Ignition-angle control, ignition 859
Ignition-point program map,
gasoline engine 650
Ignition-timing corrections, ignition 856
Ignition-voltage demand 854
Ignition-voltage demand,
gasoline engine 649
Ignition-voltage demand,
ignition coil 862
II split, antilock braking system 1267
Illuminance, lighting quantities 81
ILSAC, Lubricants 547
Image build-up,
lidar-sensor system 1824
Image display,
night vision systems 1877
Imagers, night vision systems 1878
Imbalance, wheels 1168
Immobilizer ECU,
vehicle immobilizer 1445
Immobilizer network,
vehicle immobilizer 1445
Immobilizer,
vehicle security systems 1444
Immunity against interference,
inverters 1585
Immunity to interference,
electromagnetic compatibility 1527
Impedance spectroscopy,
corrosion testing 342
Impeller,
hydrodynamic torque converter 603
Implementation of software functions,
automotive software engineering 1659
Implementations, CAN 1696
Impulse, mechanics 49
Inclination estimation,
Motorcycle Stability Control 1297
Increment sensor, rpm sensors 1740
Independence principle, tolerances 355
Independent wheel suspensions,
wheel suspensions 1150
Indian legislation, emission
legislation for motorcycles 1075
Indicated power, thermal engines 644
Indirect pulse time-of-flight method,
lidar-sensor system 1831
Indirectly measuring tire-pressure
monitoring systems 1189
Individual control,
antilock braking system 1263
Individual control, modified,
antilock braking system 1263
Individual efficiency,
thermal engines 645
Individual running resistances,
vehicle physics 464
Induced drag, aerodynamics 517
Inductance, electrical engineering 109
Induction cycle, charge cycle 665
Induction period, lubricants 539
Induction soldering, soldering 454
Induction time, diesel fuel 569
Induction voltage, ignition coil 863
Induction, electrical engineering 121
Inductive charging,
on-board charger 1601
Inductive coupling,
electromagnetic compatibility 1525
Inductive ignition, ignition 860
Inductively coupled plasma optical
emission spectrometry, elements 187
Inductor, alternators 1492
Inert-gas shielded arc welding,
welding 452

Inertia brake system 1207
Inertia weight class,
energy demand for drives 508
Inertial forces,
reciprocating-piston engine 693
Inference, sensor-data fusion 1860
In-field monitoring,
CARB legislation 1046
In-field monitoring,
emission-control legislation 1043
In-field monitoring,
EPA legislation 1050
In-field monitoring,
EU legislation 1054
Inflammation, ignition 855
Inflatable curtain,
occupant-protection systems 1421
Inflatable headrests,
occupant-protection systems 1421
Inflatable tubular system,
occupant-protection systems 1421
Inflatable tubular torso restraints,
occupant-protection systems 1421
Influencing self-steering properties
through roll stabilization 1306
Influencing self-steering
properties through variable
vibration dampers 1307
Information and warning systems,
driver-assistance systems 1896
Information areas,
display and control 1785
Information Domain Computer,
infotainment and cockpit
solutions 1781
Information recording,
traffic telematics 1802
Information transmission,
radio and TV broadcasting 1794
Information-providing parking
assistant, parking and maneuvering
systems 1884
Infotainment and cockpit solutions,
infotainment 1780
Infotainment bus, MOST 1707
Infrared headlamps,
night vision systems 1877
Infrasound, acoustics 66
Ingredients, radial tire 1171
Inhibit conditions,
on-board diagnostics 1097
Inhibitors, adhesive technologies 456
Inhibitors, coolants 554
Inhibitors, corrosion protection 346
Inhibitors, lubricants 539
Initial opening pressure,
pressure-control valve 920
Initial permeability,
material parameters 259
Initial response time, brakes 1209
Initial spring tension, natural-gas
pressure-control module 899
Injected-fuel-quantity spread,
fuel supply 821
Injection molding, plastics 322
Injection nozzles 953
Injection pressure, diesel engine 661
Injection rate, diesel engine 657
Injection ratio, thermodynamics 93
Injection-orifice alignment,
high-pressure fuel injector 850
Injection-orifice plate,
intake-manifold fuel injector 847
Injection-synchronous delivery,
high-pressure pump, common rail 939
Injector delivery compensation,
common rail 928
Injector variants,
solenoid-valve injector 930
Injectors, common rail 929
Inlet restrictor, common rail 930
Inlet valve,
antilock braking system 1265
Inlet valve, ESC 1285
In-line engine,
reciprocating-piston engine 690
In-line filter, diesel filter 922
In-line filter, gasoline filter 825
Inner disks, step-up gear units 613
Inner liner, radial tire 1171
Inner locking, locking systems 1433
Input signals,
electronic control unit (ECU) 1632
Input-power split, hybrid drives 1005
Insertion force, mechanics 50
Insertion technique,
two-component seals 410
Inside diameter, O-ring 386
Installation instructions,
radial shaft seal 403
Installation location,
12 V vehicle electrical system 1452
Installation-space section,
vehicle bodies, passenger cars 1321
Institute of Electrical and
Electronics Engineers 1648
Instrument amplifier,
operational amplifier 153
Instrument clusters, display
and control 1786

Instrumentation, display and control 1785
Insulated-gate bipolar transistor, electronics 146
Insulating foils, plastics 323
Insulating materials, plastics 322
Insulation properties, plastics 322
Insulator, spark plug 866
Intake camshaft, variable valve timing 667
Intake closes, charge cycle 665
Intake duct 842
Intake manifolds, air-intake system 763
Intake opens, charge cycle 665
Intake passage, charge cycle 665
Intake ports, 2-stroke cycle 666
Intake stroke, charge cycle 664
Intake stroke, unit injector 945
Intake valve, charge cycle 664
Intake valve, reciprocating-piston engine 685
Intake-air temperature sensor 1763
Intake-camshaft adjustment, variable valve timing 667
Intake-manifold fuel injector, mixture formation 846
Intake-manifold injection, gasoline engine 647
In-tank filter, gasoline filter 825
In-tank unit, fuel-supply module 829
Integral calculus, mathematics 215
Integral construction, vehicle bodies, commercial vehicles 1331
Integral safety, passive safety 1424
Integral, mathematics 217
Integrated charging concept, on-board charger 1601
Integrated Collision Detection Front, integral safety 1427
Integrated driving-dynamics control systems 1304
Integrated motor-generator, electrical machines 1569
Integrated power brake (IPB), brake systems for passenger cars 1226
Integrated redundancy, brake systems for passenger cars 1229
Integrated sensors, sensors 1719
Integrated starter generator, hybrid drives 995
Integration of software, automotive software engineering 1660
Integration, mathematics 216
Intelligence, artificial intelligence 255
Intelligence, sensor-data fusion 1865
Intelligent agents, artificial intelligence 249
Intelligent alternator regulation, alternators 1497
Intelligent headlamp control, driver-assistance systems 1928
Intelligent sensors, sensors 1719
Intelligent vents, occupant-protection systems 1420
Intensity curve, wave optics 79
Intensity measurement, lidar-sensor system 1824
Intensive state variables, thermodynamics 87
Interaction between molecules, chemical bonds 190
Interaction channels, display and control 1784
Intercalation electrode, batteries for electric and hybrid drives 1608
Intercity/overland buses, vehicle bodies, commercial vehicles 1330
Intercooler core, intercooling (charge-air cooling) 743
Intercooling (charge-air cooling), engine cooling 743
Interface management, vehicle bodies, passenger cars 1316
Interface regulators, alternators 1497
Interference drag, aerodynamics 517
Interference emission, electromagnetic compatibility 1525, 1531
Interference fit, rolling bearings 380
Interference immunity test, electromagnetic compatibility 1531
Interference maxima, wave optics 78
Interference minima, wave optics 78
Interference pattern, wave optics 78
Interference pulses, electromagnetic compatibility 1525
Interference suppression, electromagnetic compatibility 1534
Interference, vibrations and oscillations 61
Interference, wave optics 78
Interference-suppression measures, alternators 1500
Interframe Space, CAN 1695
Interframe symbol, Serial Wire Ring 1712
Intergranular corrosion, types of corrosion 341
Interior-monitoring systems, driver-assistance systems 1932
Interior-temperature sensor 1763

Interlocking effect, tire grip 1177
Intermediate heating, preheating systems 955
Intermediate piston, passenger-car brake system 1220
Intermediate shaft, electric final-drive unit 1028
Intermediate-speed regulation, Electronic Diesel Control 912
Internal dimensions, vehicle bodies, passenger cars 1319
Internal drag, aerodynamics 517
Internal energy, thermodynamics 87
Internal exhaust-gas recirculation, cylinder charge 814
Internal mixture formation, gasoline engine 646
Internal resistance, starter batteries 1483
Internal running resistances, energy demand for drives 507
Internal seals, pneumatic seals 406
Internal-combustion engines 642
Internal-gear pump, electric fuel pump 828
International Electrotechnical Commission 1648
Interphase reaction, corrosion processes 340
Interpretation of results, finite-element method 235
Interrupt controller, electronic control unit (ECU) 1626
Interrupt frame, Serial Wire Ring 1712
Interruption of tractive force, step-up gear units 615
Intersection assistant, driver-assistance systems 1924
In-the-loop test systems, automotive software engineering 1661
In-the-loop test 1655
Intrinsic conduction, semiconductors 139
Intrinsic photoelectric effect, sensors 1776
Intrinsically viscous flow behavior, lubricants 542
Introduction of gases, air-intake system 766
Intrusion monitoring, theft-alarm system 1449
In-use Monitor Performance Ratio, on-board diagnostics 1093
Inverse matrix, mathematics 224
Inverter control, inverters 1586
Inverters for electric drives, electrics for electric and hybrid drives 1582
Inverting amplifier, operational amplifiers 152
Involute of a circle, step-up gear units 609
Involute toothing, step-up gear units 608
Ionic bonds, chemical bonds 188
Ionic compounds, chemical bonds 189
Ionic product, acids in water 199
Ionic-current measurement procedure, spark plug 868
Ions, electrical engineering 103
IR spectroscopy, chemical bonds 191
IRHD hardness, elastomer materials 420
Iron base alloys, metallic materials 271
Iron carbide, metallic materials 271
Iron core, ignition coil 862
Iron losses, electrical machines 1579
Iron soldering, soldering 455
Iron/carbon diagram, metallic materials 271
Irradiance, lighting quantities 79
Irradiation, lighting quantities 80
Irregularity spectrum, motor-vehicle dynamics 487
Isentropic exponents, thermodynamics 91
ISO tolerance system, tolerances 355
ISO viscosity grades, lubricants 543
Isobaric change of state, thermodynamics 91
Isobutene-isoprene rubber, elastomer materials 415
Isochoric change of state, thermodynamics 91
Isolated return line, starters for heavy commercial vehicles 1517
Isolated system, thermodynamics 86
Isolating DC-DC converters 1598
Isomeric nuclides, elements 185
Isomers, elements 185
Isophone curves, acoustics 71
Isothermal change of state, thermodynamics 91
Isotopes, elements 185
IT security of AI systems, artificial intelligence 256

J

Jackknifing, driving dynamics - commercial vehicles 496
Japan Automotive Software Platform and Architecture 1648
Japanese legislation, emission legislation for motorcycles 1074
Japanese legislation, emission-control legislation for heavy commercial vehicles 1066
Japanese legislation, emission-control legislation 1054
Japanese test cycle, emission-control legislation 1062
JC08 test cycle, test cycles 1062
JE05 test cycle, test cycles for heavy commercial vehicles 1071
Joining angle, snap-on connections 448
Joining force, plug-in connections 1521
Joining force, snap-on connections 448
Joint clearance, frictional joints 433
Jointless wiper blade, windshield, rear-window and headlamp cleaning systems 1387
Joule cycle, thermodynamics 95
Joule, SI units 34
Junction-gate field-effect transistors, electronics 146

K

K factor, pressure sensors 1757
Kalman filter, sensor-data fusion 1869
Kamm circle, tires 1176
Katal, SI units 34
Kelvin, SI units 33
Kesternich, lubricants 538
Key Sharing Service, locking systems 1442
Keyed joints, frictional joints 436
Keyless entry, locking systems 1431
Kilogram, SI units 33
Kinematic viscosity 699
Kinematic viscosity, flow mechanics 57
Kinematic viscosity, lubricants 542
Kinesthetic channel, display and control 1785
Kinetic energy, mechanics 47
Kinetic energy, thermodynamics 87
Kingpin angle, chassis systems 1118
Kingpin axis, chassis systems 1118
Kingpin offset at wheel center, chassis systems 1118
Kingpin steering, steering 1192
Kirchhoff's laws, electrical engineering 110
Knee airbag, occupant-protection systems 1420
Knob, locking systems 1433
Knock control, ignition 857
Knock detection, ignition 858
Knock protection, gasoline direct injection 845
Knock sensor 1755
Knock sensor, ignition 858
Knock-control system, ignition 857
Knocking 1755
Knocking, gasoline engine 653
Knoop hardness, heat treatment 330

L

Labeling, video sensors 1856
Labyrinth system, starter batteries 1478
Ladder-type frame, vehicle bodies, commercial vehicles 1327
Lambda control, catalytic exhaust-gas treatment 879
Lambda, mixture formation, spark-ignition engines 838
Lambertian emitters, electronics 143
Laminar flow, flow mechanics 57
Laminated composite materials 303
Laminated safety glass, automotive glazing 1375
Laminates, plastics 323
Lamps with fiber-optic technology, lighting equipment 1366
Lamps with Fresnel optics, lighting equipment 1364
Lamps with light-emitting diodes, lighting equipment 1365
Lamps with reflector optics, lighting equipment 1364
Lanchester system, reciprocating-piston engine 695
Lane allocation, ACC 1892
Lane assistance, driver-assistance systems 1906
Lane centering assist, lane assistance 1908
Lane departure warning, lane assistance 1906
Lane detection, lane assistance 1906
Lane restriction, lane assistance 1910
Lane-change assistant, driver-assistance systems 1912
Lane-change assistant, lane-change assistant 1913

Lane-keeping support,
lane assistance 1908
Laplace transform, mathematics 219
Lap-trigger systems, motor sport 884
Large sunroof, sunroof systems 1406
Laser LED, electronics 143
Laser light, lighting equipment 1353
Laser technology, optical technology 82
Laser vibrometers, vehicle acoustics 532
Laser welding, welding 453
Laser-beam welding, welding 453
Laser-welded rail, common rail 943
Late fusion, sensor-data fusion 1864
Latency, vehicle communication
networks 1683
Lateral acceleration,
motor-vehicle dynamics 479
Lateral force, chassis systems 1120
Lateral forces, tires 1176
Lateral running, wheels 1168
Lateral-agility diagram,
motor-vehicle dynamics 483
Lateral-force steer, driving
dynamics - commercial vehicles 493
Law of conservation of energy,
mechanics 49
Law of mass action 197
Law of reflection, geometric optics 75
Law of refraction, geometric optics 75
Law, chemical thermodynamics 195
Laws of thermodynamics 88
Layer technology,
temperature sensor 1761
Layered structure,
deposits and coatings 350
Layout, snap-on connection 448
Lead battery, starter batteries 1478
Lead dioxide, starter batteries 1480
Lead grid, starter batteries 1478
Lead sponge, starter batteries 1479
Lead-antimony alloys,
starter batteries 1479
Lead-calcium alloys,
starter batteries 1480
Lead-calcium-tin alloys,
starter batteries 1480
Lead-in chamfers, O-ring 391
Leading axles, vehicle bodies,
commercial vehicles 1326
Leaf springs, springs 364
Leaf springs, suspension 1127
Leakage rate, seal technology 384
Lean shift, λ control 879
Lean-misfire limit, mixture formation,
spark-ignition engines 839
Learning as prediction,
artificial intelligence 250
Learning strategies,
artificial intelligence 251
Learning strategies,
control of a hybrid drive 1012
Learning, artificial intelligence 252
LED headlamps,
lighting equipment 1352
LED lights, lighting equipment 1365
LED low-beam headlamp,
lighting equipment 1352
LED, lighting equipment 1336
Left-turn assistant,
intersection assistant 1925
Legal obstacles,
future of automated driving 1936
Legal regulations, safety systems
in the motor vehicle 1414
Legal requirements,
electromagnetic compatibility 1533
Legal units, quantities and units 35
Lehr's damping ratio,
acceleration sensor 1751
Length, legal units 35
Lenses, geometric optics 76
Leveling control,
lighting equipment 1367
Lever principle, mechanics 50
Lever travel,
starters for passenger cars 1510
Lever, mechanics 50
License-plate lamps,
lighting equipment 1361
License-plate lighting,
lighting equipment 1340
Lidar sensor, lidar-sensor system 1824
Lidar-sensor system, driver assistance
and sensor system 1824
Life-cycle analysis,
energy demand for drives 514
Lifetime filter, gasoline filter 825
Lift balance, aerodynamics 520
Lift coefficient, aerodynamics 518
Lift, aerodynamics 518
Lifting device, vehicle bodies,
commercial vehicles 1326
Light beams, geometric optics 75
Light color, motor-vehicle lights
and lamps, lighting equipment 1363
Light distribution pattern,
Bi-Litronic, lighting equipment 1351
Light distribution,
lighting equipment 1350
Light functions, lighting equipment 1357

Light sources, lighting equipment 1336
Light utility vans, vehicle bodies, commercial vehicles 1325
Light/dark boundary, lighting equipment 1337, 1367
Light-alloy wheel, wheels 1160
Light-dependent resistor 1776
Light-emitting diode, electronics 142
Light-emitting diode, lighting equipment 1336, 1352
Lighting at the vehicle front end, lighting equipment 1334
Lighting at the vehicle rear end, lighting equipment 1334
Lighting equipment, lighting equipment 1335
Lighting equipment, vehicle bodies 1334
Lighting in the vehicle interior, lighting equipment 1334
Lighting quantities, optical technology 79
Lighting, display and control 1787
Light-off temperature, catalytic converter , 790, 876
Light-sensitive sensor elements 1776
Limit range, motor-vehicle dynamics 480
Limp home, diagnostics 1092
Limp-home functions, diagnostics 1092
LIN description file 1702
LIN protocol 1701
LIN, buses in motor vehicles 1700
Line cross-sections, wiring harnesses 1518
Line elements, finite-element method 235
Line protection, wiring harnesses 1520
Line routing, wiring harnesses 1520
Linear bus, vehicle communication networks 1688
Linear classifiers, sensor-data fusion 1866
Linear coefficient of thermal expansion, material parameters 258
Linear differential equations, mathematics 218
Linear momentum, mechanics 49
Linear polarization, wave optics 77
Linear press fit, frictional joints 432
Linear range, motor-vehicle dynamics 480
Linear regression, artificial intelligence 252
Linear regression, technical statistics 232
Linear single-track model, motor-vehicle dynamics 481
Linear temperature coefficient, PTC 1762
Linear-elastic behavior, mechanics 51
Liner, exhaust-gas system 796
Link capacitor, inverters 1584
Link joint, positive or form-closed joints 429
Liquefied natural gas, natural gas 580
Liquefied petroleum gas, fuels 581
Liquid cooling, batteries for electric and hybrid drives 1614
Liquid fuels, properties 586
Liquid-cooled alternator, alternators 1505
Liquid-crystal display, display and control 1789
Liquified natural gas, engines fueled by natural gas 894
Liquified petroleum gas, LPG operation 886
Liquified petroleum gas, alternative gasoline-engine operation 886
Lithium-ion cells, batteries for electric and hybrid drives 1610
Lithium-ion starter battery 1486
Lithium-ion technology, batteries for electric and hybrid drives 1608
Litronic headlamp, lighting equipment 1349
Load Adaptive Control, ESC 1281
Load capacity, rolling bearings 382
Load control, gasoline engine 656
Load frequency, tire grip 1178
Load index, tire designation 1182
Load Index, tire designation 1182
Load run, belt drives 721
Load, rolling bearings 379
Load-bearing element, vehicle bodies, commercial vehicles 1327
Load-dump protection, alternators 1494
Load-extension diagram, threaded fasteners 438
Load-point shift, electrification of the drive 974
Load-Response Driving, 12 V vehicle electrical system 1456
Load-response function, alternators 1497
Local information, traffic telematics 1802
Local Interconnect Network, buses in motor vehicles 1700

Locating/floating bearing arrangement, rolling bearings 380
Location measures, technical statistics 227
Location parameters, technical statistics 227
Lock barrel, locking systems 1433
Lock holder, locking systems 1432
Lock layout, locking systems 1434
Locking bearings, rolling bearings 379
Locking devices, threaded fasteners 443
Locking differential, differentials 628
Locking mechanism, locking systems 1432
Locking pawl, locking systems 1432
Locking systems, vehicle security systems 1430
Locking-vane pump, low-pressure diesel fuel supply 919
Lockout toothing, step-up gear units 611
Lockstep mode, electronic control unit (ECU) 1625
Lockup clutch, hydrodynamic torque converter 605
Logarithm function, mathematics 207
Logarithmic decrement, vibrations and oscillations 61
Logic bus states, CAN 1694
Long range, driver-assistance systems 1808
Longer heavier vehicles, vehicle bodies, commercial vehicles 1329
Longitudinal force, chassis systems 1120
Longitudinal forces, tires 1176
Longitudinal inclination at wheel center, chassis systems 1116
Longitudinal keyed joints, frictional joints 436
Longitudinal leaf-spring suspension, wheel suspensions 1148
Longitudinal pole, chassis systems 1116
Longitudinal tilting moment, reciprocating-piston engine 695
Longlife engine oils 540
Longlife spark plugs 867
Long-range lidar, lidar-sensor system 1827
Long-range radar, radar-sensor system 1822
Long-term goal - fully automated driving, future of automated driving 1938
Long-term loads, 12 V vehicle electrical system 1451
Loosening, threaded fasteners 443
Lorentz force, electrical engineering 125
Loss resistance, electrical engineering 128
Losses, electrical machines 1579
Lost motion, variable valve timing 670
Loudness level, acoustics 71
Loudness, acoustics 71
Low beam, Europe, lighting equipment 1341
Low beam, lighting equipment 1337
Low speed following, Adaptive Cruise Control 1894
Low-maintenance batteries, starter batteries 1483
Low-noise design, acoustics 68
Low-pin-count plug-in connections 1522
Low-pressure chamber, natural-gas pressure-control module 899
Low-pressure circuit, high-pressure pump, common rail 940
Low-pressure EGR, exhaust-gas recirculation 675
Low-pressure fuel supply, diesel engine control system 916
Low-pressure pressure-control valve, low-pressure diesel fuel supply 920
Low-pressure sensors 1758
Low-pressure system, fuel supply 822
Low-risk deployment method, occupant-protection systems 1420
Low-side switches, inverters 1583
Low-speed CAN 1693
Low-speed following, ACC 1894
Low-temperature circuit, exhaust-gas cooling 745
Low-temperature proton-exchange membrane fuel cell 1030
Low-temperature sludge, lubricants 539
Low-temperature test current, starter batteries 1483
Low-voltage ceramic glow plugs, preheating systems 956
Low-voltage metal glow plug, preheating systems 955
Low-voltage preheating systems, start-assist systems 955
LPG direct injection, LPG operation 889
LPG filler neck, LPG operation 892

LPG gaseous injector, LPG operation 891
LPG injection, LPG operation 886, 888
LPG liquid injection, LPG operation 889
LPG liquid injector, LPG operation 891
LPG operation, alternative gasoline-engine operation 886
LRR, radar-sensor system 1822
LSG panes, automotive glazing 1375
Lubricants 538
Lubricating film, friction bearings 371
Lubricating film, reciprocating-piston engine 699
Lubricating gap, reciprocating-piston engine 699
Lubricating greases 549
Lubricating oils 549
Lubrication additives, friction bearings 376
Lubricity enhancers, diesel fuel 574
Lubricity, diesel fuel 573
Lumen, SI units 34
Luminance, lighting quantities 81
Lumination, lighting quantities 81
Luminous efficacy of radiation, lighting quantities 80
Luminous efficiency, lighting quantities 81
Luminous excitance, lighting quantities 81
Luminous flux changes, lighting equipment 1350
Luminous flux, lighting quantities 81
Luminous intensities, headlamps, Europe 1342
Luminous intensities, headlamps, USA 1344
Luminous intensities, lighting equipment 1364
Luminous intensity, lighting quantities 81
Lux, SI units 34

M

“Monovalent plus” vehicles, engines fueled by natural gas 896
M+S, tire designation 1184
Machine learning techniques, video sensors 1855
Machine learning, artificial intelligence 250
Macro-roughness, tire grip 1177
Macro-scanner, lidar-sensor system 1835
Macrotick, FlexRay 1698
MAG welding, welding 452
Magnesium, nonferrous metals 280
Magnet poles, alternators 1492
Magnetic dipole moment, electrical engineering 121
Magnetic energy density, electrical engineering 121
Magnetic field quantities, electrical engineering 118
Magnetic field strength, electrical engineering 119
Magnetic field, electrical engineering 118, 125
Magnetic flux, electrical engineering 119
Magnetic induction, electrical engineering 119
Magnetic materials 290
Magnetic polarization, electrical engineering 121
Magnetic quantities, legal units 39
Magnetically inductive sensors 1726
Magnetization curves, electrical engineering 124
Magnetomotive force, alternators 1491
Magnetoresistive effects, electrical engineering 133
Magnetoresistive Random Access Memory, electronic control unit (ECU) 1628
Magnetoresistive sensors 1733
Magnetoresistors, position sensors 1729
Magnetostatic sensors 1729
Main adjustment planes, electrical seat adjustment 1408
Main box, step-up gear unit 624
Main circuit, alternators 1495
Main groups, periodic table 182
Main injection, unit injector 945
Main light functions, USA 1343
Main vent levels, climate control 1393
Main-light functions, Europe, lighting equipment 1337
Maintenance-free batteries, starter batteries 1483
Major constituents, exhaust-gas emissions 716
Malfunction indicator lamp, diagnostics 1097
Malleable cast iron, metallic materials 277
Management for spark-ignition engines 800

Maneuver emergency braking, emergency-braking systems 1919
Maneuvering assistant, parking and maneuvering systems 1886
Manganese-phosphate coatings, deposits and coatings 353
Manifold injection 841
Manifold-wall fuel-condensation effects, intake-manifold injection 842
Manipulated variable's characteristic curve, λ control 879
Manipulation possibilities, artificial intelligence 256
Manual transmission, drivetrain with internal-combustion engine 597
Manually shifted transmissions, step-up gear units 614
Map display, vehicle navigation 1875
Map matching, vehicle navigation 1873
Map-based strategies, control of a hybrid drive 1016
Map-controlled thermostat, water cooling 742
Martensite, steels 272
Martensitic structure, steels 272
Mass flow rate, flow mechanics 58
Mass locking mechanism, locking systems 1438
Mass number, elements 182
Mass spectroscopy, chemical bonds 191
Mass, legal units 36
Mass, mechanics 46
Master, vehicle communication networks 1687
Master-slave, vehicle communication networks 1687
Matching transformers, electrical engineering 129
Matching, wheels 1168
Material groups, materials 270
Material major group number, EN metallurgy standards 287
Material parameters, material parameters 258
Material selection, corrosion protection 344
Materials for direct-current relays 294
Materials for reactors 294
Materials for transformers 294
Materials 258
Mathematical signs and symbols, quantities and units 45
Mathematics 204
Matrix calculus, Mathematics 222
Matrix headlamps, lighting equipment 1358
Matrix, Mathematics 222
Maturity level, automotive software engineering 1652
Maturity levels, automotive software engineering 1652
Maximum combustion, ignition 855
Maximum current, alternators 1498
Maximum material size, tolerances 354
Maximum material virtual size, tolerances 355
Maximum permeability, material parameters 259
Maximum speed governing, Electronic Diesel Control 912
Maximum speed, alternators 1498
Maximum valve lift, reciprocating-piston engine 686
McPherson principle, wheel suspensions 1152
Mean temperature coefficient 1762
Measurement methods, radar-sensor system 1818
Measurement of diesel smoke emissions, exhaust-gas measuring techniques 1085
Measurement of reflected light, diesel smoke-emission test 1085
Measurement principles, lidar-sensor system 1828
Measurement sensitivity, acceleration sensor 1750
Measuring equipment for acoustics 532
Measuring instruments, display and control 1786
Measuring interference emission, electromagnetic compatibility 1532
Measuring interference immunity, electromagnetic compatibility 1532
Measuring noise emissions from moving vehicles, vehicle acoustics 530
Measuring noise emissions from stationary vehicles, vehicle acoustics 531
Measuring tool, automotive software engineering 1662
Mechanical centrifugal supercharger, mechanical superchargers 772
Mechanical efficiency, thermal engines 644
Mechanical fully variable valve gear 669
Mechanical stress-measuring systems, acceleration sensor 1752

Mechanical-technological material parameters, material parameters 260
Mechanics, basic principles 46
Mechatronic systems 1640
Mechatronics 1640
Media Oriented Systems Transport, buses in motor vehicles 1707
Media resistance, elastomer materials 421
Median, technical statistics 227
Medium- and heavy-duty trucks, vehicle bodies, commercial vehicles 1325
Medium range, driver-assistance systems 1808
Medium-pressure sensors 1758
Melt-blown fibers, air-intake system 757
Melting temperature, material parameters 258
Memory function, electrical seat adjustment 1408
ME-Motronic 805, 806
Mesh topology, vehicle communication networks 1690
Meshing, belt drives 728
Message Authentication Code, vehicle immobilizer 1447
Message format, CAN 1695
Message identifier, vehicle communication networks 1688
Message scheduling, LIN 1702
Message-oriented method, vehicle communication networks 1688
Messages, vehicle communication networks 1682
Metal active-gas welding, welding 452
Metal electrode, batteries for electric and hybrid drives 1607
Metal hydride, batteries for electric and hybrid drives 1607
Metal inert-gas welding, welding 452
Metal injection molding, sintered metals 280
Metal monolith, catalytic converter 792
Metal powder sintering, sintered metals 280
Metal soaps, lubricants 540, 549
Metal springs, springs 364
Metal-bead gaskets, flat seals 395
Metal-ceramic bearings, friction bearings 375
Metal-film resistor, temperature sensor 1762
Metal-film temperature resistor 1762
Metal-free carbon coatings, deposits and coatings 352
Metallic bond, metallic materials 271
Metallic bonds, chemical bonds 190
Metallic materials, materials 271
Metallic paint, deposits and coatings 351
Metastable form, substances 192
Meter, SI units 32
Metering module, exhaust-gas treatment, diesel engines 968
Metering precision, common rail 932
Metering unit, common rail 939
Methanol, alcohol operation 901
Methanol, gasoline 569
Metric fine thread, threaded fasteners 445
Metric ISO thread, threaded fasteners 445
Metric standard thread, threaded fasteners 445
Microbuses, vehicle bodies, commercial vehicles 1330
Microcomputer, electronic control unit (ECU) 1624
Microcontroller unit, electronic control unit (ECU) 1625
Microcontrollers, electronic control unit (ECU) 1626
Micro-Doppler effect, active pedestrian protection 1921
Micromechanical bulk silicon acceleration sensors 1753
Micromechanical pressure sensors 1757
Micromechanical yaw-rate sensors 1743
Micromechanics 1720
Microporous separators, starter batteries 1478
Microprocessor, electronic control unit (ECU) 1624
Micro-roughness, tire grip 1177
Micro-scanner, lidar-sensor system 1836
Mid fusion, sensor-data fusion 1864
Mid-engine drive, drivetrain with internal-combustion engine 599
Midibuses, vehicle bodies, commercial vehicles 1330
Midpoint viscosity, lubricants 542
Mid-range lidar, lidar-sensor system 1827
Mid-range radar, radar-sensor system 1822
Midside nodes, finite-element method 235

MIG welding, welding 452
Mild hybrid,
electrification of the drive 971
Mild hybrid, hybrid drives 996
Mileage,
commercial-vehicle tires 1173
Milestones in automobile history,
vehicle bodies, passenger cars 1310
Mineral oils, lubricants 540
Mineral-oil fluids, brake fluids 559
Minibuses, vehicle bodies,
commercial vehicles 1330
Minimum clamping force,
threaded fasteners 439
Minimum noise from vehicles,
vehicle acoustics 531
Minimum tread depth, tire tread 1174
Minimum wrap angles,
belt drives 724, 728
Minimum yield point,
threaded fasteners 438
Minor constituents,
exhaust-gas emissions 716
Misalignment, radar sensor 1893
Missing electrons,
electrical engineering 103
Mixed friction, lubricants 543
Mixer, exhaust-gas system 791
Mixture formation,
diesel engine 657, 658
Mixture formation,
gasoline engine 646, 647, 648
Mixture formation, LPG operation 889
Mixture formation,
spark-ignition engines 838
Mixture-controlled combustion,
diesel engine 659
Mixture-formation systems,
spark-ignition engines 840
M-Motronic, management
for spark-ignition engines 805
MNEDC, test cycles 1059
Modal analysis 64
Model library, mechatronics 1643
Model predictive control,
control of a hybrid drive 1012
Model wind tunnel 527
Model-based development, automotive
software engineering 1654
Models of E/E architecture, architecture
of electronic systems 1674
Models, mechatronics 1642
Modifications, substances 192
Modified service life, rolling
bearings 383
Modulation methods,
lidar-sensor system 1829
Module, batteries for electric
and hybrid drives 1612
Modulus of elasticity,
material parameters 260
Modulus of elasticity, materials 270
Modulus of elasticity, mechanics 51
Molar volume,
substance concentrations 194
Molar weight,
substance concentrations 194
Mold flow analysis,
two-component seals 408
Mold module, inverters 1591
Molded seal, seals 393
Mole percent,
substance concentrations 194
Mole, SI units 33
Mole, substance concentrations 194
Molecular adhesion, tire grip 1178
Molecular orbitals, chemical bonds 190
Molecular spectroscopy,
chemical bonds 191
Molybdenum disulfide, lubricants 540
Moment of ignition, ignition 855
Moment of inertia, mechanics 46, 48
Moment, mechanics 46
Moment-of-ignition determination,
ignition 861
Monitoring algorithms,
diagnostics 1090
Monitoring functions,
ESC, commercial vehicles 1291
Monitoring in driving mode,
diagnostics 1090
Monitoring module, electronic
control unit (ECU) 1630
Monitoring system, ESC 1286
Monitoring, CAN 1696
Monitoring, Motronic 804
Monocrystalline silicon semiconductor
resistors, temperature sensors 1763
Monolithic integrated circuits,
electronics 157
Monomode fiber,
optical fibers/waveguides 83
Monomolecular reaction, elements 187
Monovalent vehicles,
engines fueled by natural gas 894
MOS field-effect transistors,
electronics 146
MOST 150 1707
MOST 25 1707
MOST 50 1707

MOST application layer 1709
MOST frame structure 1708
MOST, buses in motor vehicles 1707
Motive force,
motor-vehicle dynamics 471
Motor Industry Software Reliability Association 1648
Motor mode, hybrid drives 993
Motor octane number, gasoline 565
Motor sport 882
Motorcycle ABS,
Motorcycle Stability Control 1298
Motorcycle battery,
starter batteries 1486
Motorcycle Stability Control,
chassis control 1296
Motorcycle Traction Control,
Motorcycle Stability Control 1302
Motorcycle,
antilock braking system 1268
Motorcycles,
overall motor-vehicle system 29
Motor-gear unit, windshield,
rear-window and headlamp cleaning systems 1381
Motor-sport connectors,
motor sport 883
Motor-sport spark plugs,
motor sport 885
Motor-vehicle bulbs,
lighting equipment 1337
Motor-vehicle dynamics,
vehicle physics 470
Motronic versions for motorcycles 809
Motronic 801
Mounting elements,
exhaust-gas system 796
Mounting mat, catalytic converter 793
Mounting sleeve, O-ring 391
Movable magnet,
position sensors 1732
Moving-off information system for heavy commercial vehicles, information and warning systems 1902
MPC, control of a hybrid drive 1012
MRR, radar-sensor system 1822
Mud and Snow, tire designation 1184
Mufflers, exhaust-gas system 794
Multibeam lidar,
lidar-sensor system 1835
Multi-camera system, information and warning systems 1897
Multicast, vehicle communication networks 1688
Multi-circuit brake system 1207
Multi-color injection molding,
two-component seals 408
Multi-component injection molding,
two-component seals 408
Multi-cone synchromesh,
step-up gear units 612
Multi-disk brake, step-up gear units 613
Multi-disk clutch,
step-up gear units 613
Multifuel engines,
internal-combustion engines 662
Multifunction regulators,
alternators 1497
Multigrade oils 540
Multi-hole valve,
high-pressure fuel injector 850
Multilayer friction bearing 372
Multilink independent wheel suspension,
wheel suspensions 1152
Multi-master, vehicle
communication networks 1687
Multimedia channel, MOST 1707
Multimedia data,
transmission via MOST 1708
Multimode fibers,
optical fibers/waveguides 84
Multipath reception, receivers 1799
Multi-phase converters,
DC-DC converters 1596
Multiphase system,
three-phase current system 178
Multiple-circuit protection valve, brake systems, commercial vehicles 1236
Multiple-compartment lamp,
lighting equipment 1340
Multipole wheel, rpm sensors 1739
Multispark ignition, ignition coil 864
Multistage carburizing,
heat treatment 337
Multi-step parking, parking
and maneuvering systems 1884
Muscular-energy brake system 1207
Muscular-energy steering system,
steering 1193
Mushroom-head rivet, riveting 458

N

Nanofiber-based filter media,
air-intake system 770
Narrow-band antennas,
electrical engineering 129
Narrow-band interference,
electromagnetic compatibility 1526
Narrow-band spectrum, acoustics 68

Natural constants,
quantities and units 44
Natural frequency, modal analysis 64
Natural gas, fuels 579
Natural oscillation 61
Natural radioactivity, elements 186
Natural-gas direct injection 897
Natural-gas injector 897
Natural-gas manifold injection 896
Natural-gas pressure-control
module 898
Natural-gas vehicles 894
Natural-oscillation characteristics,
modal analysis 64
Natural-oscillation shape,
modal analysis 64
Navigation devices 1872
Navigation systems 1872
NCAP assessment,
active pedestrian protection 1921
NCAP assessment,
emergency-braking systems 1918
NCAP, safety systems
in the motor vehicle 1414
NDIR analyzer, exhaust-gas
measuring techniques 1082
NDUV analyzer, exhaust-gas
measuring techniques 1084
NdYAG lasers, laser welding 453
Near-field conditions,
electrical engineering 129
Near-field lidar, lidar-sensor system 1827
Near-infrared systems,
night vision systems 1877
Needle bearing, rolling bearings 378
Néel point, magnetic materials 290
Negative electrodes,
starter batteries 1478
Negative feedback,
operational amplifiers 150
Negative grids, starter batteries 1479
Neper, legal units 39
Nernst cell, λ sensors 1766
Nernst equation, chemistry 203
Net calorific value, fuels 562
Network layer, vehicle
communication networks 1685
Network management, LIN 1703
Network master, MOST 1709
Network model, architecture
of electronic systems 1677
Network topology, vehicle
communication networks 1688
Neural networks,
control of a hybrid drive 1016
Neural networks,
sensor-data fusion 1866
Neutral-steering,
motor-vehicle dynamics 482
Neutrons, elements 182
New car assessment program, safety
systems in the motor vehicle 1414
New planning,
artificial intelligence 255
New York City Cycle, test cycles 1059
Newton meter, legal units 37
Newton, SI units 34
Newtonian fluids, lubricants 542
Nickel deposits,
deposits and coatings 349
Nickel-metal hydride technology,
batteries for electric and hybrid
drives 1607
Nickel-silver, nonferrous metals 277
Nighttime vision, lighting quantities 81
Night-vision systems 1876
NiMH cell, batteries for electric
and hybrid drives 1607
Nitric oxides,
exhaust-gas emissions 717
Nitriding hardness depth,
heat treatment 338
Nitriding steel, metallic materials 276
Nitriding, heat treatment 338
Nitrile butadiene rubber,
elastomer materials 416
Nitrocarburizing, heat treatment 338
Nitrogen oxides,
exhaust-gas emissions 717
Nitrogen oxides, gasoline engine 655
Nitrogen-oxide emissions,
diesel engine 661
Nitrous gases,
exhaust-gas emissions 717
NLGI grade, lubricating greases 549
NMOS transistors, electronics 148
Noble metals, chemistry 203
Noble-gas configuration,
chemical bonds 189
Node address, vehicle
communication networks 1688
Node model, architecture
of electronic systems 1676
Noise emission from vehicles,
vehicle acoustics 530
Noise emissions from stationary
vehicles, vehicle acoustics 531
Noise immissions, acoustics 70
Noise reduction, acoustics 68, 70
Noise, operational amplifiers 155

No-load-current consumers, 12 V vehicle electrical system 1451
No-load-current management, 12 V vehicle electrical system 1463
Nominal capacity, starter batteries 1482
Nominal slip values, ESC 1277
Non Return to Zero, buses in motor vehicles 1692
Non Return to Zero, vehicle communication networks 1685
Nonconductors, electrical engineering 110
Non-destructive access control, CAN 1695
Non-electrochemical corrosion-testing procedures, corrosion testing 343
Nonferrous metals, metallic materials 277
Nonferrous-metal alloys, EN metallurgy standards 288
Non-inverting amplifier, operational amplifiers 152
Non-isolating high-output DC-DC converters 1595
Nonlinearities, control engineering 245
Non-locking bearings, rolling bearings 379
Nonmetallic inorganic materials 303
Non-muscular-energy brake system 1207
Non-oriented electrical steel strip 292
Non-return valve, ESC 1285
Nonwovens, engine lubrication 752
Normal concentration, chemistry 202
Normal hydrogen electrode, chemistry 202
Normal potential, chemistry 202
Normalizing, heat treatment 336
Not finally annealed electrical steel strip 292
NO_x sensor 1772
NO_x storage catalytic converter, catalytic exhaust-gas treatment 875
NO_x storage catalytic converter, exhaust-gas treatment, diesel engines 966
NO_x storage, NO_x storage catalytic converter 875
Nozzle needle, piezo inline injector 934
Nozzle-holder assembly 953
NRZ process, buses in motor vehicles 1692
NSC method, exhaust-gas treatment, diesel engines 966
NSC regeneration, exhaust-gas treatment, diesel engines 966
NTC thermistor 1760
NTC, temperature sensor 1760
Nuclear charge number, elements 182
Nuclear constituents, elements 182
Nuclear fission, elements 186
Nuclides, elements 185
Null frame indicator, FlexRay 1699
Number systems, mathematics 204
Numbering the cylinders, reciprocating-piston engine 691
Numbers, mathematics 204
Numerical aperture, optical fibers/waveguides 84
Numerical modal analysis 65
Nusselt number, thermodynamics 99

O

OAT radiator protection agents, coolants 555
OBD functions, on-board diagnostics 1098
OBD I, on-board diagnostics 1092
OBD II, on-board diagnostics 1092
OBD requirements for heavy commercial vehicles 1101
OBD threshold values, on-board diagnostics 1094
Object class, driver-assistance systems 1809
Object classification, driver-assistance systems 1899
Object detection, ACC 1892
Object detection, driver-assistance systems 1899
Object detection, sensor-data fusion 1861, 1866
Object prediction, lane-change assistant 1913
Object-model approaches, sensor-data fusion 1862
Obstacles to automated driving, future of automated driving 1935
Occupant sensing, occupant-protection systems 1422
Occupant-protection systems, passive safety 1416
Octane number, gasoline 564
Octave band spectrum, acoustics 68
Octet rule, chemical bonds 189
Offline approaches, control of a hybrid drive 1012
Offline calibration, automotive software engineering 1663

Off-road cars,
overall motor-vehicle system 29
Off-road vehicles,
overall motor-vehicle system 29
Offset spray angle,
intake-manifold fuel injector 848
Offset voltage,
operational amplifier 151, 154
Offset,
reciprocating-piston engine 691
Off-state saturation current,
sensors 1777
Ohm, SI units 34
Ohm's law, electrical engineering 110
Ohmic resistance,
electrical engineering 110
Oil change interval,
engine lubrication 752
Oil cooler,
reciprocating-piston engine 688
Oil cooling, engine cooling 745
Oil film,
reciprocating-piston engine 699
Oil filter,
reciprocating-piston engine 678
Oil filters, engine lubrication 751
Oil pan,
reciprocating-piston engine 682
Oil pressure,
reciprocating-piston engine 700
Oil pump,
reciprocating-piston engine 678, 688
Oil reservoir,
reciprocating-piston engine 688
Oil supply,
reciprocating-piston engine 688
Oil temperature,
reciprocating-piston engine 699
Oil throughput, lubricants 539
Oil wedge,
reciprocating-piston engine 699
Oil-cooler functions 682
Oil-film theory,
reciprocating-piston engine 701
Oil-pump housing,
reciprocating-piston engine 682
Oil-retaining volume,
reciprocating-piston engine 699
Oil-to-air coolers, oil cooling 745
Oil-to-coolant coolers, oil cooling 746
Oil-to-water heat exchanger,
engine lubrication 753
Oil-volume flow, shock absorbers 1137
OLED, display and control 1790
OLED, electronics 143
Omega recess,
reciprocating-piston engine 676
On-board charger, electrics
for electric and hybrid drives 1600
On-Board Diagnostics for
motorcycles 1105
On-board diagnostics 1092
One-battery vehicle electrical system,
12 V vehicle electrical system 1458
One-box design, vehicle bodies,
passenger cars 1316
One-pedal control,
control of electric drives 986
One-plunger radial-piston pump,
common rail 939
One-step parking, parking and
maneuvering systems 1884
Online approaches, control
of a hybrid drive 1012
Online calibration, automotive
software engineering 1663
Opacimeter,
diesel smoke-emission test 1085
Opacity measurement,
diesel smoke-emission test 1085
Open cycles, thermal engines 643
Open deck,
reciprocating-piston engine 682
Open loop, driving-dynamics test
procedures 498
Open system, thermodynamics 86
Open systems and their interfaces for
electronics in motor vehicles 1648
Open vessels, hydrostatics 56
Opening chain, locking systems 1432
Opening in the partition,
starter batteries 1478
Opening phase,
solenoid-valve injector 933
Opening pressure,
pressure-control valve 920
Opening times,
reciprocating-piston engine 686
Open-loop control,
control engineering 242
Open-loop gain,
operational amplifier 151
Open-loop-controlled adaptation,
control engineering 246
Operating characteristics,
synchronous machine 170
Operating conditions,
electronic control unit (ECU) 1621
Operating conditions, sensors 1717
Operating Data, Motronic 804

Operating forces,
threaded fasteners 439
Operating strategies,
gasoline engine 647
Operating strategies,
starter batteries 1477
Operating strategy, diesel engine 662
Operating strategy, vehicle electrical systems for hybrid and electric vehicles 1565
Operating temperature,
catalytic converter 876
Operating unit,
display and control 1788
Operating-point frequency,
electrical machines 1568
Operating-temperature range,
elastomer materials 415
Operational amplifier, electronics 149
Operational performance,
electrical machines 1574
Opposed-cylinder (boxer) engine 690
Opposed-piston engine 690
Optical fibers/waveguides,
lighting equipment 1366
Optical fibers/waveguides,
optical technology 83
Optical technology 74
Optimization-based operating strategy,
control of a hybrid drive 1012
Optoelectronic sensors 1776
Orbital energies, periodic table 185
Orbitals, periodic table 185
Organic light-emitting diode,
electronics 143
Organic light-emitting diodes,
display and control 1790
O-ring, seals 386
Oscillation excitation,
electrical machines 1568
Oscillation gyrometers 1743
OSI reference model, vehicle communication networks 1684
Other alloyed steels,
EN metallurgy standards 286
Otto cycle, thermodynamics 94
Out of plane, yaw-rate sensor 1744
Out of Range Check,
on-board diagnostics 1100
Outer casing, radial shaft seal 399
Outer disks, step-up gear units 613
Outer locking, locking systems 1433
Outlet restrictor, common rail 930
Outlet valve,
antilock braking system 1266
Outlet valve, ESC 1286
Out-of-position detection,
occupant-protection systems 1423
Output resistance,
operational amplifier 151
Output-power split, hybrid drives 1005
Outside-temperature sensor 1763
Oval clearance, friction bearings 371
Oval countersunk-head rivet,
riveting 458
Overall efficiency,
thermal engines 644, 645
Overall motor-vehicle system,
motor vehicles 28
Overdrive, drivetrain with internal-combustion engine 597
Overexcitation,
synchronous machine 171
Overflow pressure regulator,
fuel supply, spark-ignition engine 835
Overflow restriction,
low-pressure diesel fuel supply 920
Overflow valve, high-pressure pump,
common rail 940
Overflow valves,
four-circuit protection valve 1235
Overhead bucket-tappet assembly,
reciprocating-piston engine 686
Overlap control,
transmission control 632
Overlap TDC, charge cycle 664
Overlapping joint,
adhesive technologies 456
Overload operation,
electrical machines 1567
Overmodulation, receivers 1799
Overrunning clutch,
starters for passenger cars 1510
Oversteering,
motor-vehicle dynamics 482
Overstressing, threaded fasteners 441
Overturning limit, ESC 1289
Overvoltage protection, inverters 1584
Owner key, locking systems 1442
Oxidation equations,
three-way catalytic converter 874
Oxidation stability, diesel fuel 573
Oxidation, chemistry 202
Oxygen concentration sensor,
chemistry 203
Oxygen storage,
three-way catalytic converter 875
Oxygen supply, fuel cell 1036
Oxygen zero transition point,
λ control 879

Oxygenates, fuels 562
Oxymethylene ethers, fuels 584
Ozone resistance,
elastomer materials 422

P

p quantile, technical statistics 227
P0 topology, hybrid drives 1000
P1 topology, hybrid drives 1000
P2 topology, hybrid drives 1001
P3 topology, hybrid drives 1002
P4 topology, hybrid drives 1002
Package data channel, MOST 1707
Package data,
transmission via MOST 1708
Packets, vehicle communication
networks 1682
Paint coatings,
deposits and coatings 350
Paints, deposits and coatings 350
Palladium, catalytic converter 874
Panel and roller pressboard,
plastics 323
Panel van, vehicle bodies,
commercial vehicles 1324
Panhard rod,
wheel suspensions 1148
Panorama roofs,
automotive glazing 1378
Panoramic sunroof,
sunroof systems 1406
Parabola, mathematics 205
Paraboloids, geometric optics 77
Paraffinic diesel fuels 577
Paraffins, fuels 560
Parallel combinations, springs 363
parallel hybrid, hybrid drives 998
Parallel parking, parking and
maneuvering systems 1884
Parallel starting system, starters for
heavy commercial vehicles 1514
Parallelogram of forces, mechanics 49
Paramagnetic detector, exhaust-gas
measuring techniques 1083
Paramagnetic materials,
electrical engineering 122
Paramagnets, magnetic materials 290
Parameter, automotive
software engineering 1663
Parity bit, Serial Wire Ring 1713
Parity bit, vehicle communication
networks 1691
Parking aids, parking and
maneuvering systems 1882
Parking and maneuvering systems,
driver-assistance systems 1882
Parking brake system 1206
Parking brake, automatic
brake-system operations 1294
Parking brake, disk brake 1251
Parking brake, drum brake 1259
Parking brake,
passenger-car brake system 1222
Parking lamps, lighting equipment 1361
Parking lock,
electric final-drive unit 1028
Parking lock, step-up gear units 619
Parking maneuver assistant, parking
and maneuvering systems 1888
Parking position, windshield,
rear-window and headlamp
cleaning systems 1383
Parking-brake system, brake systems
for passenger cars and light utility
vehicles 1216
Parking-brake system, brake systems,
commercial vehicles 1232
Parking-brake valve, brake systems,
commercial vehicles 1238
Part-flash lidar,
lidar-sensor system 1835
Partial charges, chemical bonds 190
Partial flow meter,
hot-film air-mass meters 1749
Partial integration, mathematics 217
Partial lift,
piezo high-pressure fuel injector 853
Partial state of charge,
12 V vehicle electrical system 1464
Partial-engine operation,
cylinder shutoff 818
Partially active control, ESC 1285
Partially loaded systems,
suspension systems 1132
Particle composite materials 303
Particle filters, cabin filters 1397
Particle velocity 60
Particle velocity, acoustics 66
Particulate counting, exhaust-gas
measuring techniques 1084
Particulate emission, exhaust-gas
measuring techniques 1083
Particulate emissions,
diesel engine 661
Particulate filter 793
Particulate filter, catalytic exhaust-gas
treatment, gasoline engine 881
Particulate filter, exhaust-gas
treatment, diesel engines 963

Particulate sensor 1773
Particulate size distribution, exhaust-gas measuring techniques 1084
Particulate-filter diagnosis, on-board diagnostics 1100
Particulate-filter process, exhaust-gas measuring techniques 1083
Pascal, SI units 34
Passenger cars, overall motor-vehicle system 29
Passenger-car air filter 755
Passenger-car air-intake system 755
Passenger-car intake manifolds 763
Passenger-car tires 1173
Passenger-compartment climate control 1392
Passing (overtaking), motor-vehicle dynamics 476
Passing distance, motor-vehicle dynamics 477
Passivation, micromechanics 1721
Passive control, ESC 1285
Passive entry, locking systems 1431
Passive rpm sensors 1738
Passive safety functions, driver-assistance systems 1806
Passive safety, safety systems in the motor vehicle 1413
Passive safety, vehicle bodies, commercial vehicles 1332
Passive star, vehicle communication networks 1689
Passive vibration isolation 64
Patch-array-antenna system, radar-sensor system 1821
Path, mechanics 46
PAX system, wheels 1158, 1167
Payload length, FlexRay 1699
Payload preamble indicator, FlexRay 1699
Payload, FlexRay 1699
Payload, vehicle communication networks 1688
PDP method, exhaust-gas measuring techniques 1080
Peak factor, vibrations and oscillations 62
Pearlite, steels 272
Pearlitic structure, steels 272
Pedal characteristic, electrohydraulic brake 1224
Pedal-feel simulator, brake systems for passenger cars 1227
Pedal-travel sensor, electrohydraulic brake 1224
Pedal-travel sensor, electronically controlled brake system 1242
Pedal-travel simulator, electrohydraulic brake 1224
Pedestrian protection, emergency-braking systems 1920
Pedestrian protection, occupant-protection systems 1422
Peltier coefficient, electrical engineering 132
Peltier effect, electrical engineering 132
Pencil coil, ignition coil 863
Penetration clinching, penetration-clinching processes 459
Penetration, lubricants 541
Penetration-clinching process, permanent joints 459
Pentagram connection, alternators 1500
Pentahedrons, finite-element method 236
Perceived sound levels, acoustics 71
Percentile, technical statistics 227
Perception of the environment, sensor-data fusion 1860
Perception, lidar-sensor system 1825
Perception, sensor-data fusion 1860
Perceptron, sensor-data fusion 1866
Perfect engine, internal-combustion engines 644
Perfectly Keyless, locking systems 1441
Performance characteristic, electric final-drive unit 1026
Performance classes, electric final-drive unit 1026
Performance specs, mechatronics 1644
Performant ECUs, architecture of electronic systems 1673
Period 60
Periodic emission testing, EU legislation 1054
Periodic meshing, belt drives 728
Periodic table, chemistry 182
Periodical course of irregularity, motor-vehicle dynamics 487
Peripheral Sensor Interface, buses in motor vehicles 1705
Permanent joints, joining and bonding techniques 450
Permanent-field synchronous machine, electrical machines 1573
Permanent-magnet materials 299
Permanent-magnet materials, electrical engineering 123

Permeability, material parameters 259
Permeation limits, emission legislation for motorcycles 1072
Perpendicular parking, parking and maneuvering systems 1885
Perpetuum mobile, thermodynamics 88
PES headlamps, lighting equipment 1348
PFC rectifier, on-board charger 1604
pH value, chemistry 199
pH value, strong acids 201
pH value, weak acids 201
Phase voltage, inverters 1583
Phase-Change Memory, electronic control unit (ECU) 1628
Phase-in, CARB legislation 1045
Phase-in, EPA legislation 1049
Phases, substances 192
Phase-shifted full-bridge, DC-DC converters 1598
Phenol-formaldehyde resins, plastics 322
Phon, acoustics 71
Phosphating coatings, deposits and coatings 353
Photocells 1777
Photocurrent, sensors 1777
Photodiode, electronics 142
Photodiodes 1777
Photoelectric effect, sensors 1776
Photoelectric reflected-light barrier, sensors 1778
Photometric characteristic quantities, lighting quantities 79
Photometric quantities, legal units 39
Photometrical terms and definitions, lighting equipment 1345
Photometry, lighting quantities 79
Photon energy 1776
Photonic-crystal fiber, optical fibers/waveguides 83
Photoresistors 1776
Phototransistors 1777
Photovoltaic effect, sensors 1777
Physical characteristic quantities, lighting quantities 79
Physical layer, vehicle communication networks 1684
Physical material parameters 258
Physical vapor deposition, deposits and coatings 352
Pi filter, inverters 1585
Piezo actuator, piezo high-pressure fuel injector 851
Piezo high-pressure fuel injector, mixture formation 851
Piezo inline injector, common rail 933
Piezoceramic element, knock sensor 1755
Piezoelectric acceleration sensors 1753
Piezoelectric effect, acceleration sensor 1752
Piezoelectric knock sensor 1755
Pilot control, transmission control 640
Pilot injection, axial-piston distributor pump 950
Pilot injection, unit injector 945
Pin connections, positive or form-closed joints 428
Pin-fin concept, inverters 1592
Pinhole-camera model, video sensors 1847
Pipe threads, threaded fasteners 446
Piston bearing surface, reciprocating-piston engine 682
Piston clearance, reciprocating-piston engine 712
Piston guidance, rod and piston sealing systems 413
Piston normal force, reciprocating-piston engine 692
Piston recess, reciprocating-piston engine 676
Piston ring, reciprocating-piston engine 677, 701
Piston seal, O-ring 391
Piston seals, pneumatic seals 406
Piston seals, seals 413
Piston shapes, reciprocating-piston engine 676, 678
Piston speed, reciprocating-piston engine 713
Piston stroke, reciprocating-piston engine 706
Piston, reciprocating-piston engine 676, 684
Piston-pin force, reciprocating-piston engine 692
Piston-ring shapes, reciprocating-piston engine 679
Piston-surface geometry, reciprocating-piston engine 676
Piston-travel function, reciprocating-piston engine 692
Pitch angle, threaded fasteners 437
Pitch angle, vehicle motion 463
Pitch axis, chassis systems 1118
Pitch axis, yaw-rate sensor 1743
Pitch, acoustics 71

Pitch, belt drives 727
Pitch, threaded fasteners 437
Pitch, vehicle motion 463
Pitch-circle diameter, wheels 1156
Pitching moment, vehicle motion 463
Pitot-tube flow meters, sensors 1746
Pitting corrosion, types of corrosion 341
PK profile, belt drives 722
Plain bearings 371
Planar technology, electronics 157
Planck's quantum of action 83
Plane of oscillation, wave optics 78
Plane system of forces, mechanics 49
Plane triangle, mathematics 210
Plane wave,
vibrations and oscillations 62
Planetary gear,
starters for passenger cars 1508
Planetary gears, step-up gear units 610
Planetary-gear set, steering 1201
Planetary-gear set,
step-up gear units 610
Planning, artificial intelligence 254
Plasma discharge,
lighting equipment 1348
Plasma, substances 192
Plastic flow behavior, lubricants 542
Plastic wheel, wheels 1160
Plasticity, control of a hybrid drive 1018
Plastics, materials 306
Plate capacitor,
electrical engineering 108
Plate straps, starter batteries 1478
Plate, finite-element method 237
Plate-cone measuring system,
lubricants 541
Platform software, automotive
software engineering 1647
Platinum resistors,
temperature sensor 1762
Platinum, catalytic converter 874
Plausibility check, diagnostics 1091
Plenum chamber, air-intake system 763
Plenum chamber,
intercooling (charge-air cooling) 743
Plug-in connections 1520
Plug-in connections,
electronic control unit (ECU) 1636
Plug-in hybrid,
electrification of the drive 971
Plug-in hybrid, hybrid drives 998
Plunger-and-barrel assembly,
electric fuel pump 827
PMOS transistors, electronics 148
pn junction, semiconductor 139
PN PEMS, exhaust-gas
measuring techniques 1084
Pneumatic actuators,
air-intake system 766
Pneumatic brake assistant, automatic
brake-system operations 1292
Pneumatic seals, Seals 406
Point cloud, lidar-sensor system 1825
Point of interest,
vehicle navigation 1874
Poisson distribution,
technical statistics 228
Poisson's ratio,
material parameters 260
Polar coordinates, mathematics 211
Polar substances, lubricants 541
Polarization curve, fuel cell 1032
Polarization, electrical engineering 117
Polarization, wave optics 77
Polarization-resistance measurement,
corrosion testing 342
Pollutant dispersion system,
wind tunnels 527
Pollutant emissions, diesel engine 661
Pollutant emissions,
gasoline engine 655
Pollutant formation, diesel engine 661
Pollutant formation,
gasoline engine 655
Pollutant reduction inside the engine,
diesel engine 661
Pollutant reduction, diesel engine 661
Pollutant reduction,
gasoline engine 655
Pollutants, exhaust-gas emissions 716
Pollutants, gasoline engine 655
Poly ellipsoid system,
lighting equipment 1348
Polyacrylic rubber,
elastomer materials 416
Polyester resins, plastics 319
Polyester, plastics 325
Polygon effect, chain drives 732
Polygon ring, high-pressure pump,
common rail 938
Polymer bearings, friction bearings 376
Polynomial, mathematics 205
Polytetrafluoroethylene, lubricants 541
Polytropic curve change of state,
thermodynamics 91
Polyurethane adhesive,
adhesive technologies 456
Polyurethane, plastics 325
Polyurethanes, elastomer materials 417
Poly-α-olefins, lubricants 543

Poor-weather light, lighting equipment 1356
Porosity, material parameters 262
Porous lead dioxide, starter batteries 1479
Porous silicon, micromechanics 1722
Portable emission measuring devices, exhaust-gas measuring techniques 1084
Position detection, electrical seat adjustment 1408
Position determination, GPS 1872
Position lamps, lighting equipment 1360
Position sensors for transmission control 1731
Position sensors 1724
Positional tolerance, tolerances 356
Position-controlled acceleration sensor 1751
Positioning, vehicle navigation 1872
Position-measuring systems, acceleration sensors 1750
Positive belt drives, internal-combustion engines 725
Positive feedback, operational amplifiers 150
Positive grids, starter batteries 1479
Positive or form-closed joints, detachable connections 426
Positive-displacement pump, electric fuel pump 827
Positive-displacement supercharger, mechanical superchargers 773
Positive-locking grip, tire grip 1178
Positron, elements 186
Post-heating phase, preheating systems 955
Post-ignitions, spark plug 868
Postprocessor, finite-element method 234
Potential energy, mechanics 47
Potential energy, thermodynamics 87
Potential-free housing, starters for heavy commercial vehicles 1517
Potentially susceptible devices, electromagnetic compatibility 1526
Potentiometric measurement methods, hydrogen sensors 1775
Potentiostat, corrosion testing 342
Pour point, lubricants 541
Powder composite materials, magnetic materials 294
Power balance, lidar-sensor system 1828
Power coupling procedure, electromagnetic compatibility 1531
Power curve, supercharging 672
Power Density Line, alternators 1505
Power diodes, electronics 141
Power in the alternating-current circuit, electrical engineering 126
Power output per liter, reciprocating-piston engine 703
Power smoothing, on-board charger 1604
Power supply unit, motor sport 884
Power switches, inverters 1583
Power take-up elements, drivetrain 600
Power yield, gasoline engine 656
Power, legal units 37
Power, mechanics 49
Power-assisted steering system, steering 1193
Power-assisted steering systems for commercial vehicles, steering 1203
Power-assisted steering systems for passenger cars, steering 1196
Power-assisted steering systems with modulation capability, steering 1196
Power-factor correction, on-board charger 1603
Power-neutral downsizing, hybrid drives 993
Power-socket consumption, energy demand for drives 511
Power-split hybrid, hybrid drives 1005
Power-steering pump, steering 1198
Power-steering system, steering 1193
Power-sunroof drives, sunroof systems 1406
Power-supply rejection ratio, operational amplifier 151
Power-window systems 1404
Prandtl number, thermodynamics 99
Preamble, Ethernet 1704
Prebraking time, motor-vehicle dynamics 475
Prechamber, wind tunnels 525
Precipitation hardening, heat treatment 336
Pre-crash detection, safety systems in the motor vehicle 1411
Pre-Crash Positioning, integral safety 1427
Pre-engaged-drive starter, starters for passenger cars 1509
Pre-excitation circuit, alternators 1494
Prefilling, automatic brake-system operations 1294

Prefix symbol, SI units 33
Preheating phases,
preheating systems 955
Preheating software,
preheating systems 955
Preheating systems,
start-assist systems 954
Preheating time,
flame start system 960
Pre-heating, preheating systems 955
Pre-ignition, ignition 859
Pre-ignition, spark plug 868, 871
Preliminary detent,
locking systems 1432
Preload forces, threaded fasteners 441
Preloaded bearing arrangement,
rolling bearings 380
Premium lock, locking systems 1434
Premixed combustion,
diesel engine 658
Preprocessor,
finite-element method 234
Press fit, frictional joints 430
Press-fit force, frictional joints 432
Pressure build-up time, brakes 1209
Pressure build-up time,
motor-vehicle dynamics 476
Pressure control, common rail 926
Pressure generation, common rail 926
Pressure increase, gasoline engine 651
Pressure modulator,
electrohydraulic brake 1225
Pressure on the face of a hole,
mechanics 51
Pressure ratio, thermodynamics 95
Pressure reduction, natural-gas
pressure-control module 898
Pressure regulator, brake systems,
commercial vehicles 1234
Pressure sensors 1756
Pressure tank, fuel supply 825
Pressure vessel, flow mechanics 58
Pressure vs. volumetric-flow map,
supercharging 673
Pressure-backing, unit injector 945
Pressure-compensated valve,
common rail 931
Pressure-control module,
engines fueled by natural gas 898
Pressure-control modules, electronically
controlled brake system 1245
Pressure-control valve,
gasoline direct injection 833
Pressure-control valve,
low-pressure diesel fuel supply 920
Pressure-control valve,
traction control system 1270
Pressure-control valve,
transmission control 637
Pressure-drop phase,
antilock braking system 1261
Pressure-increase ratio,
thermodynamics 93
Pressure-indicated cycle power,
thermal engines 644
Pressure-loaded ball valve,
common rail 930
Pressure-relief valve, natural-gas
pressure-control module 899
Pressure-retention phase,
antilock braking system 1261
Pressure-rise phase,
antilock braking system 1261
Pressure-wave superchargers,
turbochargers and
superchargers 772
Pressure-wave supercharging 673
Prestress, threaded fasteners 438
Prestroke, unit injector 945
Pretensioning force, belt drives 721
Primary address,
Serial Wire Ring 1713
Primary detent, locking systems 1432
Primary quantum number,
periodic table 185
Primary retarder, continuous-operation
brake systems 1247
Primary shoe, drum brake 1257
Primary winding, ignition coil 862
Principal maxima, wave optics 79
Priority control, vehicle
communication networks 1687
Prismatic hardcase cell, batteries
for electric and hybrid drives 1611
Prisms, geometric optics 76
Process assessment, automotive
software engineering 1653
Process capability characteristic value,
technical statistics 233
Process capability,
technical statistics 233
Process description models, automotive
software engineering 1650
Process evaluation models, automotive
software engineering 1651
Process variables, thermodynamics 86
Process, thermodynamics 86
Processing methods, plastics 322
Product Specifications, tolerance 354
Production date, tire designation 1183

Profiled shaft-hub connections, positive or form-closed joints 427
Program memory, electronic control unit (ECU) 1629
Projection system, headlamps, lighting equipment 1346
Prometheus, driver-assistance systems 1804
Prompted troubleshooting, ECU diagnostics 1106
Propagation of sound, acoustics 67
Properties, finite-element method 237
Property classes, threaded fasteners 438
Proportional solenoid, transmission control 639
Protection against misuse, starters for heavy commercial vehicles 1517
Protocol, MOST 1708
Protocols, vehicle communication networks 1682
Protolysis, acids in water 198
Protons, elements 182
Pseudo-first-order reaction, reaction kinetics 197
Pseudo-Hall sensors 1735
PSI5, buses in motor vehicles 1705
Psychoacoustics, vehicle acoustics 535
PTC thermistor 1760
PTC, temperature sensor 1760
PTFE, lubricants 541
Pull-in winding, starters for passenger cars 1508
Pulsation, acoustics 70
Pulse detection, lidar-sensor system 1830
Pulse supercharging, exhaust-gas turbocharger 780
Pulse time-of-flight method, lidar-sensor system 1829
Pulse wheel, rpm sensors 1738
Pulses in the vehicle electrical system, electromagnetic capability 1525
Pump cell, λ sensors 1767
Punch riveting, riveting 458
Puncture detection, tire-pressure monitoring systems 1188
Purely electric driving, hybrid drives 993
Push voltage, preheating systems 955
Pushbelt, step-up gear unit 622
Push-rod piston, passenger-car brake system 1220
PVD coatings, deposits and coatings 352
PWM activation, DC-DC converters 1598
PWM inverter, vehicle electrical systems for hybrid and electric vehicles 1561
PWM valves, transmission control 636
PWR inverters, inverters 1582
Pyrometer 1760

Q

Quad-headlamp system, lighting equipment 1340
Quadratic temperature coefficient, PTC 1762
Quality assurance, automotive software engineering 1654
Quality control, gasoline engine 656
Quality factor, vibrations and oscillations 61
Quantities used in atom physics, legal units 40
Quantities, quantities and units 32
Quantity control, gasoline engine 656
Quantity of heat, legal units 37
Quantity of light, lighting quantities 81
Quantum numbers, periodic table 182
Quarter-vehicle model, motor-vehicle dynamics 488
Quartile, Technical statistics 227
Quenching and drawing, heat treatment 335
Quenching, spark plug 869
Queue, Serial Wire Ring 1714

R

“Restored-signal recognition”, diagnostics 1091
R12, refrigerants 591
R-1234yf, climate control 1396
R1234yf, refrigerants 592
R134a, climate control 1396
R134a, refrigerants 591
R-744, climate control 1396
Rack travel, chassis systems 1119
Rack, steering 1195
Rack-and-pinion steering, steering 1195
Radar-antenna radiation pattern, lane-change assistant 1913
Radar-sensor system, driver assistance and sensor system 1816
Radial crushing strength, material parameters 262
Radial engine, reciprocating-piston engine 690

Radial play, rolling bearings 379
Radial run-out, radial shaft seal 401
Radial seal, plug-in connections 1521
Radial shaft seal, seals 398
Radial tire 1171
Radial vee-form filter,
gasoline filter 826
Radial-piston distributor pumps 951
Radial-piston pump,
gasoline direct injection 832
Radian, SI units 34
Radiance, lighting quantities 80
Radians, mathematics 207
Radiant energy, lighting quantities 79
Radiant excitance,
lighting quantities 80
Radiant flux, lighting quantities 79
Radiant intensity, lighting quantities 80
Radiation quantities,
lighting quantities 79
Radiation thermometer 1760
Radiator core, water cooling 737
Radiator design, water cooling 742
Radiator designs, water cooling 737
Radiator protection agents
containing silicate, coolants 554
Radiator protection agents,
coolants 553
Radiator tank, water cooling 738
Radio and TV reception 1794
Radio Data System,
radio and TV broadcasting 1799
Radio Data System,
traffic telematics 1801
Radio interference,
radio and TV interference 1796
Radio tuners 1797
Radioactive decay, elements 186
Radioactivity, elements 186
Radiocarbon method 185
Radiometric characteristic quantities,
lighting quantities 79
Radiometry, lighting quantities 79
Radionuclides, elements 185
Rail, common rail 943
Rail, fuel supply, gasoline engine 833
Rail-pressure sensor 1759
Rain sensor 1778
Rain sensor, windshield,
rear-window and headlamp
cleaning systems 1388
RAM,
electronic control unit (ECU) 1629
Raman spectroscopy,
chemical bonds 191
Ram-tube supercharging,
air-intake system 763
Random Access Memory,
electronic control unit (ECU) 1629
Random variable,
technical statistics 230
Range extender,
electrification of the drive 971
Range extender, hybrid drives 998
Range, radar-sensor system 1822
Range-change box,
step-up gear unit 624
Range-extender operation,
control of electric drives 989
Ranges, lidar-sensor system 1827
Ranging sensor, ACC 1890
Rapid control prototyping, automotive
software engineering 1657
Rare earths, periodic table 185
Rated current,
12 V vehicle electrical system 1455
Rated current, alternators 1498
Rated power,
reciprocating-piston engine 705
Rated speed, alternators 1498
Rating sound level, acoustics 70
Rational numbers, mathematics 204
Rationality Check,
on-board diagnostics 1100
Rattling noises, vehicle acoustics 535
Ravigneaux gear set,
step-up gear units 620
Raw-air intake, air-intake system 768
Raw-air line, air-intake system 755
Ray model, optical technology 74
RDE test, test cycles 1061
RDS/TMC standard,
traffic telematics 1801
RE gas,
engines fueled by natural gas 894
Reaction kinetics,
reactions of substances 196
Reaction progression,
chemical thermodynamics 195
Reaction time, brakes 1209
Reaction time,
motor-vehicle dynamics 475
Reactions of substances 195
Reactive power,
electrical engineering 127
Reactive-power compensation,
electrical engineering 127
Readiness codes,
on-board diagnostics 1097
Real gas, thermodynamics 90

Real numbers, mathematics 204
Real Road PEMS Tests, test cycles for heavy commercial vehicles 1070
Real-time capable, software 1646
Rear cross-traffic warning, information and warning systems 1900
Rear fog warning lamps, lighting equipment 1362
Rear reflectors, lighting equipment 1361
Rear underride guard, vehicle bodies, commercial vehicles 1333
Rear wiper drives, windshield, rear-window and headlamp cleaning systems 1389
Rear wiper systems, windshield, rear-window and headlamp cleaning systems 1388
Rear-axle steering, steering 1192
Rear-end collision, integral safety 1426
Rear-engine drive, drivetrain with internal-combustion engine 599
Rearview camera, parking and maneuvering systems 1886
Rearview-camera system, information and warning systems 1896
Rear-wheel drive, drivetrain with internal-combustion engine 598
Rear-wheel lift-up control, Motorcycle Stability Control 1298
Rebound resilience, elastomer materials 422
Rebound stage, shock absorbers 1141
Reception problems, radio and TV broadcasting 1795
Recessive state, CAN 1692
Reciprocating-piston engine types 690
Reciprocating-piston engine 676
Recirculated-air mode, climate control 1392
Recirculating-ball gear, steering 1200
Recirculating-ball steering, steering 1196
Recoil permeability, material parameters 259
Recording of air charge, cylinder charge 815
Recrystallization annealing, heat treatment 335
Rectification value, vibrations and oscillations 62
Rectification, alternators 1493
Rectifier diode, electronics 141
Rectilinear motion, mechanics 46
Recuperation level, control of electric drives 986
Recuperation torque, control of electric drives 986
Recuperation, 12 V vehicle electrical system 1464
Recuperation, alternators 1497
Recuperation, hybrid drives 992
Redox reaction, chemistry 203
Reduced stress, mechanics 52
Reduction equations, catalytic converter 874
Reduction gear, starters for passenger cars 1508
Reduction gear, steering 1199
Reduction, chemistry 202
Redundancy, brake systems for passenger cars 1228
Reference belt length, belt drives 722
Reference mark, rpm sensors 1741
Reference variable, control engineering 242
Reference-vacuum chamber, pressure sensors 1758
Reflectance factor, electrical engineering 128
Reflected image, night vision systems 1877
Reflection coefficient, geometric optics 76
Reflection headlamps, lighting equipment 1347
Reflection mufflers, exhaust-gas system 794
Reflection system, headlamps, lighting equipment 1346
Reflector focal length, lighting equipment 1346
Reflector illuminated area, lighting equipment 1346
Reflectors, geometric optics 77
Refractive index, geometric optics 75, 76
Refrigerant circuit, climate control 1395
Refrigerant cooling, batteries for electric and hybrid drives 1614
Refrigerant, climate control 1395, 1396
Refrigerants, operating fluids 591
Refueling emissions, evaporative-emissions test 1089
Refueling test, evaporative-emissions test 1089
Regeneration of the particulate filter, catalytic exhaust-gas treatment, gasoline engine 881

Regeneration of the particulate filter, exhaust-gas treatment, diesel engines 963
Regeneration, NO_x storage catalytic converter 875
Regeneration, particulate filter 793
Regeneration-air tank, brake systems, commercial vehicles 1234
Regenerative braking system, electrification of drive 976
Regenerative braking system, electrification of the drive 972
Regional programs, emission-control legislation for heavy commercial vehicles 1066
Regroovable, tire tread 1174
Regulated preheating system, start-assist system 957
Regulation of climate control 1394
Reinforcement additives, friction bearings 376
Reinforcement learning, artificial intelligence 251
Relative frequency, technical statistics 227
Relative permeability, electrical engineering 121
Relative-pressure sensors 1756
Relaxation behavior, plastics 309
Relaxation, springs 364
Relay-station attack, locking systems 1431
Release angle, snap-on connections 448
Release force, snap-on connections 448
Release of ammonia, exhaust-gas treatment, diesel engines 969
Release torque, threaded fasteners 441
Reliability, sensors 1717
Remagnetization losses, electrical engineering 124
Remanence point, electrical engineering 122
Remanence, electrical engineering 122
Remote frame, CAN 1695
Remote key, locking systems 1431
Remote welding, welding 453
Remote-controlled parking assistant, parking and maneuvering systems 1888
Renewable-energy gas, engines fueled by natural gas 894
Repeater, vehicle communication networks 1689
Replacement of wheel-torque distribution by brake and engine-control interventions 1306
Replacing the battery, starter batteries 1489
Replay attack, vehicle immobilizer 1446
Representative sinor, inverters 1587
Requirement management, architecture of electronic systems 1678
Requirement model, architecture of electronic systems 1675
Research octane number, gasoline 565
Reset function, tire-pressure monitoring systems 1189
Residual exhaust gas, cylinder charge 812
Residual gas, charge cycle 664
Residual ripple, electromagnetic compatibility 1525
Residual speed, wet braking 1175
Resilient clips, snap-on connections 447
Resilient snap-on hooks, snap-on connections 447
Resistance measurement 1761
Resistance projection welding, welding 450
Resistance spot welding, welding 450
Resistance to flow, flow mechanics 59
Resistance to seizing, friction bearings 372
Resistance to wear, friction bearings 372
Resistance welding, welding 450
Resistance, electrical engineering 106
Resistance-based measurement principle, hydrogen sensors 1775
Resistivity, electrical engineering 106
Resolving capacity, wave optics 79
Resonance chamber, air-intake system 764
Resonance flap, air-intake system 764
Resonance plenum, air-intake system 764
Resonance sharpness, acceleration sensor 1751
Resonance sharpness, vibrations and oscillations 61
Resonance, vibrations and oscillations 61
Resonant frequency, acceleration sensor 1751
Resonant frequency, vibrations and oscillations 61
Resonator, laser technology 83

Resonator-type mufflers,
air-intake system 758
Response calculation,
modal analysis 65
Response model,
blind-spot assistant 1913
Response model,
emergency-braking systems 1917
Response model,
lane-change assistant 1913
Response model, modal analysis 65
Response time,
motor-vehicle dynamics 475
Retarder, continuous-operation
brake systems 1247
Retention rate, charge cycle 664
Retention systems,
occupant-protection systems 1418
Retreadable,
commercial-vehicle tires 1173
Retreading,
commercial-vehicle tires 1173
Return pump,
antilock braking system 1265
Return pump, ESC 1285
Returnless system, fuel supply 820
Reverberation rooms,
vehicle acoustics 532
Reverse assembly,
pressure sensors 1757
Reverse automatic emergency braking,
emergency-braking systems 1922
Reverse-current block, alternators 1494
Reversible adiabatic change of state,
thermodynamics 91
Reversing drives, windshield,
rear-window and headlamp cleaning
systems 1383
Reversing lamps,
lighting equipment 1362
Reversing technology, windshield,
rear-window and headlamp cleaning
systems 1384
Reynolds number, flow mechanics 57
Reynolds number, thermodynamics 99
RFID transponders,
vehicle immobilizer 1445
Rheological shock-absorber systems,
shock absorbers 1139
Rheology, lubricants 541
Rheopexy, lubricants 542
Rhodium, catalytic converter 874
Ribbed V-belt profile, belt drives 722
Ribbed V-belt, belt drives 721
Rich shift, λ control 879
Ride comfort segments,
vehicle acoustics 532
Ride comfort simulation,
vehicle acoustics 532
Ride comfort, chassis systems 1110
Ride down benefit,
occupant-protection systems 1420
Ride-height control,
chassis systems 1132
Riding physics,
Motorcycle Stability Control 1296
Rigid axles, wheel suspensions 1148
Rim base, wheels 1156
Rim bead seat, wheels 1155
Rim circumference, wheels 1156
Rim designs, wheels 1157
Rim diameter, wheels 1156
Rim flange, wheels 1155
Rim offset, wheels 1156
Rim size, wheels 1157
Rim, wheels 1155
Ring gear,
starters for passenger cars 1509
Ring gear, step-up gear units 610
Ring structure, MOST 1709
Ring topology, vehicle
communication networks 1689
Ring-shaped snap-on connections 447
Ripple in vehicle electrical system,
electromagnetic compatibility 1525
RISC processor,
electronic control unit (ECU) 1624
Rise path, electrical engineering 122
Rivet nut, riveting 458
Riveting, permanent joints 457
Road excitation,
motor-vehicle dynamics 486
Road illumination,
lighting equipment 1357
Road-sign recognition,
driver-assistance systems 1899
Road-speed limitation,
Electronic Diesel Control 913
Roadworks assistant,
lane assistance 1909
Robustness, control engineering 245
Rocker arm,
reciprocating-piston engine 685
Rockwell hardness, heat treatment 328
Rod guides,
rod and piston sealing systems 412
Rod seal, O-ring 392
Rod seal, pneumatic seals 406
Rod seals, seals 411
Rod shape, rpm sensors 1739

Roll angle, vehicle motion 463
Roll axis, chassis systems 1116
Roll behavior,
motor-vehicle dynamics 485
Roll pole, chassis systems 1116
Roll steer, driving dynamics - commercial vehicles 493
Roll, vehicle motion 463
Roll-back effect, disk brake 1251
Roller electrodes, seam welding 451
Roller race,
starters for passenger cars 1510
Roller sun blind, sunroof systems 1406
Roller tappet,
gasoline direct injection 830
Roller-cell pump, electric fuel pump 827
Roller-type chain, chain drives 730
Roller-type overrunning clutch,
starters for passenger cars 1510
Rolling bearings, machine parts 378
Rolling elements, rolling bearings 378
Rolling friction, mechanics 54
Rolling moment, vehicle motion 463
Rolling pairs,
reciprocating-piston engine 701
Rolling radius, chassis systems 1121
Rolling resistance,
running resistances 464
Rolling resistance, tires 1180
Rolling-contact pressure,
reciprocating-piston engine 701
Rolling-resistance coefficient,
tires 1180
Rollover Mitigation Function, ESC 1281
Rollover prevention, ESC 1281
Rollover protection,
occupant-protection systems 1421
Root function, mathematics 205
Roots supercharger,
mechanical superchargers 773
Rope friction, mechanics 54
Rotary drive, windshield,
rear-window and headlamp cleaning systems 1384
Rotary drives, windshield,
rear-window and headlamp cleaning systems 1382
Rotary motion, mechanics 46
Rotary oscillator, yaw-rate sensor 1744
Rotary seal for tire-pressure control,
tires 1190
Rotary table, two-component seals 409
Rotary-magnet actuator,
distributor injection pump 949
Rotating load, rolling bearings 381
Rotating mass, mechanics 47
Rotating-field generation,
three-phase current system 179
Rotational impulse, mechanics 49
Rotational motions, vehicle motion 463
Rotational speed, mechanics 46
Rotocap,
reciprocating-piston engine 686
Rotor shapes, rpm sensors 1739
Rotor, alternators 1492
Roughness values,
reciprocating-piston engine 701
Round filters, air-intake system 770
Route computation,
traffic telematics 1802
Route computation,
vehicle navigation 1874
Route guidance,
vehicle navigation 1875
Route recommendations,
vehicle navigation 1874
Route-guidance systems,
traffic telematics 1802
Routing, vehicle communication networks 1685
rpm sensors 1738
Rubber compound, belt drives 721, 726
Rubber compounds, plastics 316
Rubber elasticity, plastics 316
Rubber elements,
suspension systems 1124
Rubber friction, tires 1179
Rubber lip, windshield, rear-window and headlamp cleaning systems 1387
Rubber springs, suspension 1126
Rubber, plastics 316
Run length, chain drives 733
Run-flat tires 1171
Run-in performance,
friction bearings 372
Running resistance, drivetrain with internal-combustion engine 596
Running resistances,
vehicle physics 464
Running-loss test,
evaporative-emissions test 1088
Run-time environment, automotive software engineering 1649
Rust level, corrosion testing 343

S

Sacrificial anode,
corrosion protection 345
Sacrificial layer, micromechanics 1721

SAE viscosity grades,
lubricants 543, 547
Safe function, locking systems 1438
Safety element, air-intake system 770
Safety functions, ESC,
commercial vehicles 1291
Safety functions, locking systems 1438
Safety labyrinth cover,
starter batteries 1483
Safety margin,
motor-vehicle dynamics 476
Safety requirements, batteries
for electric and hybrid drives 1611
Safety strategy, transmission control 634
Sailing operation, hybrid drives 1001
Salts, chemical bonds 189
Sample size, technical statistics 227
Sample, technical statistics 228
Sanding waves,
vibrations and oscillations 62
Sandwich injection molding,
two-component seals 408
Satellite positioning system GPS 1872
Saturation detection, catalytic exhaust-
gas treatment, gasoline engine 881
Saturation detection, exhaust-gas
treatment, diesel engines 964
Saturation polarization,
electrical engineering 122
SC03 cycle, test cycles 1058
Scalability, electric final-drive unit 1027
Scalar product, mathematics 214
Scan tool,
on-board diagnostics 1093, 1097
Scavenge efficiency, charge cycle 664
Scavenging pumps, 2-stroke cycle 666
Scavenging, cylinder charge 814
Scavenging,
engines fueled by natural gas 897
Schematic diagram 1546
Schottky diode, electronics 142
Scleroscope hardness,
heat treatment 331
SCR method, exhaust-gas treatment,
diesel engines 966
Screw/bolt, threaded fasteners 437
Screw-type supercharger,
mechanical superchargers 774
Scrub radius, chassis systems 1119
Seal seat, spark plug 867
Seal technology, machine parts 384
Sealed-beam design,
lighting equipment 1344
Seal-glass bonding,
micromechanics 1722
Sealing caps, seals 396
Sealing compounds, plastics 323
Sealing compounds, properties 324
Sealing cup, radial shaft seal 399
Sealing gap, molded seal 394
Sealing gap, O-ring 387
Sealing geometry,
natural-gas injector 898
Sealing lips, plug-in connections 1521
Sealing-edge rings,
rod and piston sealing systems 411
Seals, machine parts 384
Seam welding, welding 451
Seat adjustment,
electrical seat adjustment 1408
Seat Belt Reminder, safety systems
in the motor vehicle 1410
Seat belts,
occupant-protection systems 1418
Seat reference point,
vehicle bodies, passenger cars 1318
Seat-occupancy mat,
occupant-protection systems 1422
Secant modulus,
snap-on connections 448
Second law, thermodynamics 88
Second, SI units 32
Secondary brake system 1206
Secondary braking effect, brake
systems, commercial vehicles 1230
Secondary Collision Mitigation,
occupant-protection systems 1417
Secondary groups, periodic table 182
Secondary hardening,
heat treatment 334
Secondary lock,
plug-in connections 1522
Secondary maxima, wave optics 79
Secondary retarder, continuous-
operation brake systems 1248
Secondary shoe, drum brake 1257
Secondary winding, ignition coil 862
Secondary-air injection,
catalytic-converter heating 877
Secondary-air pump,
catalytic-converter heating 878
Secondary-air valve,
catalytic-converter heating 878
Secondary-brake system, brake
systems for passenger cars and
light utility vehicles 1217
Secondary-brake system, brake
systems, commercial vehicles 1232
Second-order reaction,
reaction kinetics 197

Section designation, circuit diagrams 1549
Security, vehicle immobilizer 1446
Seebeck coefficient, electrical engineering 131
Seebeck effect, electrical engineering 131
Segment sensor, rpm sensors 1740
Seiliger cycle, thermodynamics 93
Seismic mass, acceleration sensor 1753
Seismic mass, knock sensor 1755
Seizing, reciprocating-piston engine 699
Selection criteria, rolling bearings 379
Selective catalytic reduction, exhaust-gas treatment, diesel engines 966
Select-low control, antilock braking system 1263
Self-aligning rolling bearing, rolling bearings 378
Self-augmenting effect, drum brake 1257
Self-caused dirtying 523
Self-cleaning temperature, spark plug 867
Self-discharge, starter batteries 1483, 1487
Self-excitation, 12 V vehicle electrical system 1455
Self-excitation, alternators 1491
Self-induction, electrical engineering 121
Self-inhibiting effect, drum brake 1257
Self-lubricating bearings made of metal 374
Self-lubricating bearings made of plastic 376
Self-protection function, electronic control unit (ECU) 1635
Self-stabilization, Motorcycle Stability Control 1296
Self-steering effect, motor-vehicle dynamics 481
Self-test, diagnostics 1090
Semiconductor devices, electronics 141
Semiconductor memories, electronic control unit (ECU) 1628
Semiconductor technology, electronics 138
Semiconductor technology, micromechanics 1720
Semiconductors, electrical engineering 110
Semicrystalline thermoplastics, plastics 307
Semi-finished, electrical steel strip 292
Semi-hollow rivet, riveting 458
Semi-independent axles, wheel suspensions 1149
Semi-rigid axles, wheel suspensions 1148
Semi-synthetic filter media, air-intake system 757
Semi-synthetic oils, lubricants 544
Sensor classification, sensors 1718
Sensor flap, flow meters 1746
Sensor redundancy concept, architecture of electronic systems 1672
Sensor shapes, rpm sensors 1739
Sensor space, sensor-data fusion 1860
Sensor-data fusion, ACC 1894
Sensor-data fusion, driver assistance and sensor system 1860
Sensor-data fusion, driver-assistance systems 1809
Sensors, automotive electronics 1716
Sensotronic brake control, passenger-car brake system 1224
Separate motor-generator, electrical machines 1571
Separators, starter batteries 1478
Sequence control in the master, Serial Wire Ring 1714
Sequential supercharging, complex supercharging systems 785
Serial testing, emission-control legislation 1043
Serial Wire Ring, buses in the motor vehicle 1711
Series combinations, springs 363
Series hybrid, hybrid drives 1003
Series-based ECUs, motor sport 884
Series-parallel hybrid, hybrid drives 1005
Service brake system 1206
Service Information System, diagnostics 1106
Service-brake system, brake systems for passenger cars and light utility vehicles 1216
Service-brake system, commercial vehicles 1230
Service-brake system, tractor vehicle 1230
Service-brake system, trailer 1231
Service-brake valve, brake systems, commercial vehicles 1237

Service-brake valve, electronically controlled brake system 1244
Servo unit, steering 1199
Servomotor, steering 1199
Setpoint time gap, ACC 1891
Settling chamber, wind tunnels 525
SFTP cycles, test cycles 1058
Shaft, radial shaft seal 401
Shafts, axles and shafts 359
Shakers, vehicle acoustics 532
Sharpness, acoustics 71
SHARX, receivers 1800
Shear layer, aerodynamics 522
Shear modulus curve, plastics 309
Shear rate, lubricants 541
Shear stress, flow mechanics 57
Shear stress, lubricants 541
Shear stress, material parameters 262
Shear stress, mechanics 52
Shear viscosity, lubricants 541
SHED, evaporative-emissions test 1087
Shedding, aerodynamics 517
Shedding, starter batteries 1484
Shell elements, finite-element method 235
Shell model, finite-element method 237
Shell-type mufflers, exhaust-gas system 795
Shift-point control, transmission control 632
Shock-absorber characteristic curves, shock absorbers 1138
Shock-absorber piston, shock absorbers 1136
Shore hardness, elastomer materials 419
Short-circuit scavenging, 2-stroke cycle 666
Short-circuiting-ring sensor 1726
Short-range radar, radar-sensor system 1822
Short-term loads, 12 V vehicle electrical system 1451
Short-time performance, electric final-drive unit 1026
Shoulder-belt pretensioner, occupant-protection systems 1418
Shrink fit, frictional joints 433
Shrinking process, spark plug 866
Shutoff valve, LPG operation 893
SI units, quantities and units 32
Side airbag, occupant-protection systems 1421
Side electrode, spark plug 867
Side force, aerodynamics 518
Side member, vehicle bodies, commercial vehicles 1328
Side-marker lamps, lighting equipment 1361
Side-view assistant, lane-change assistant 1912
Sidewall markings, radial tire 1172
Sidewall, radial tire 1171
Siemens, SI units 34
Sievert, SI units 34
Sign function, mathematics 206
Signal conditioning, electronic control unit (ECU) 1632
Signal lamps, lighting equipment 1359
Signal path, diagnostics 1091
Signal processing, electronic control unit (ECU) 1634
Signal Range Check, on-board diagnostics 1100
Signal transmission, radio and TV broadcasting 1794
Signal types, electronic control unit (ECU) 1634
Signal-to-noise ratio, radio and TV broadcasting 1796
Silica, starter batteries 1478
Silica, tires 1181
Silicon semiconductor resistors 1763
Silicone adhesive, adhesive technologies 456
Silicone rubber, elastomer materials 417
Silicone-oil fluids, brake fluids 559
Silicones, plastics 325
Simple lock, locking systems 1434
Simple regenerative braking system, electrification of drive 976
Simplex brake with S-cam, drum brakes 1258
Simplex brake, drum brake 1257
Simpson gear set, step-up gear units 620
Simulation, mechatronics 1642
Sine function, mathematics 208
Single box, air-intake system 767
Single clutch, power take-up elements 600
Single ring, vehicle communication networks 1689
Single rocker-arm assembly, reciprocating-piston engine 685
Single spray, intake-manifold fuel injector 848
Single-barrel pump, gasoline direct injection 830

Single-beam lidar,
lidar-sensor system 1834
Single-circuit brake system 1207
Single-circuit version, steering 1203
Single-component adhesives,
adhesive technologies 456
Single-core seal,
plug-in connections 1522
Single-disk clutch,
power take-up elements 600
Single-ECU concept,
LPG operation 890
Single-ECU concepts,
engines fueled by natural gas 896
Single-flow ventilation, alternators 1504
Single-line brake system 1208
Single-pane toughened safety glass,
automotive glazing 1375
Single-phase substances,
chemistry 192
Single-spark ignition coil,
ignition coil 863
Single-track vehicle model, ESC 1277
Single-tube shock absorbers,
shock absorbers 1137
Single-wire cable, CAN 1693
Sinter filter, natural-gas
pressure-control module 899
Sintered bearings, friction bearings 374
Sintered metals for soft-magnetic
components 294
Sintered metals, metallic materials 280
Sintered-ceramic resistors,
temperature sensors 1761
Sinusoidal steering input, driving-
dynamics test procedures 501
Sipes, tire tread 1174
Situation development,
intersection assistant 1924
Size, sensors 1717
Skin friction drag, aerodynamics 516
Skyhook control strategy,
shock absorbers 1142
Skyhook shock absorber,
shock absorbers 1142
Slack run, belt drives 721
Slack run, chain drives 734
Slave address, Serial Wire Ring 1713
Slave, vehicle communication
networks 1687
Sleep mode, LIN 1703
Sleeve-/tooth-type chain,
chain drives 731
Sleeve-type chain, chain drives 731
Slew rate, operational amplifiers 155
Sliding bearings 700
Sliding fit,
positive or form-closed joints 427
Sliding friction, mechanics 53
Sliding shoes, drum brake 1257
Sliding/titling sunroof,
sunroof systems 1406
Slip angle, chassis systems 1120
Slip angle, tires 1176
Slip control, ESC 1278
Slip stiffness, driving dynamics -
commercial vehicles 494
Slip, chassis systems 1121
Slip, tires 1176
Slip, wheel-slip control systems 1260
Slip-switching threshold,
antilock braking system 1262
Sloping teeth,
starters for passenger cars 1510
Slot, FlexRay 1698
Slot, vehicle communication
networks 1686
Slow response, OBD functions 1100
Slug-down,
electronic control unit (ECU) 1637
Slug-up,
electronic control unit (ECU) 1637
Small-area contrast,
lighting quantities 82
Small-signal range,
motor-vehicle dynamics 479
Smartphone connection, infotainment
and cockpit solutions 1782
Smoke number,
diesel smoke-emission test 1086
Smoke tester,
diesel smoke-emission test 1086
Smooth-running control,
common rail 928
Smooth-running control,
Electronic Diesel Control 913
Snap-on connections,
detachable connections 447
Snell's law of refraction,
geometric optics 75
SOC window, vehicle electrical systems
for hybrid and electric vehicles 1564
Society of Automotive Engineers 1648
Soft annealing, heat treatment 335
Soft magnetic materials 290
Soft magnetic materials,
electrical engineering 123
Soft soldering, soldering 454
Soft stop, electrohydraulic brake 1225
Softeners, radial tire 1171

Soft-magnetic ferrite cores 294
Soft-material seals, flat seals 395
Software architecture 1647
Software components, automotive software engineering 1648
Software development, automotive software engineering 1650, 1654
Software model, architecture of electronic systems 1677
Software Process Improvement and Capability Determination, automotive software engineering 1653
Software, electronic control unit (ECU) 1638
Solar cell, electronics 142
Solar gas, engines fueled by natural gas 894
Solder, soldering 454
Soldering, permanent joints 454
Solenoid switch, starters for passenger cars 1508
Solenoid-valve injectors, common rail 929
Solid angle, legal units 35
Solid angle, lighting quantities 82
Solid mesh, finite-element method 237
Solid modifications, substances 192
Solid polymer bearings, friction bearings 376
Solid rivet, riveting 458
Solid, finite-element method 237
Solid-body friction, lubricants 543
Solid-color paint, deposits and coatings 351
Solid-state lidar sensors, lidar-sensor system 1834
Solubility product, law of mass action 198
Sommerfeld number, friction bearings 371
Soot emissions, diesel engine 661
Sound absorption coefficient, acoustics 68
Sound absorption, acoustics 68
Sound absorption, exhaust-gas system 791
Sound cleaning, vehicle acoustics 535
Sound damping, acoustics 68
Sound design, vehicle acoustics 535
Sound engineering, vehicle acoustics 535
Sound exposure level, acoustics 70
Sound field, acoustics 69
Sound identification, tire designation 1185
Sound insulation, acoustics 68
Sound intensity level, acoustics 69
Sound intensity, acoustics 67
Sound power level, acoustics 69
Sound power, acoustics 67, 69
Sound pressure level, acoustics 69
Sound pressure, acoustics 66
Sound quality, vehicle acoustics 535
Sound spectrum, acoustics 68
Sound, acoustics 66
Sound-field quantities, acoustics 69
Sources of interference, electromagnetic compatibility 1525
Sources of noise, vehicle acoustics 533
Space vector, inverters 1587
Space-charge region, diode 140
Spacing of transmission, drivetrain with internal-combustion engine 597
Spacing, step-up gear units 615
Spagnolet, locking systems 1432
Spark current, ignition coil 863
Spark duration, ignition 854
Spark plug, Ignition 865
Spark position, spark plug 869
Spark voltage, ignition coil 863
Spark-plug choice, spark plug 868
Spark-plug concepts, spark plug 869
Spark-plug electrodes, gasoline engine 649
Spark-plug faces, spark plug 873
Spark-plug installation, spark plug 872
Spark-plug practice, spark plug 872
Spark-plug shell, spark plug 866
Spark-plug versions, spark plug 871
Specific fuel consumption, energy demand for drives 506
Specific heat capacity, material parameters 258
Specific heat, material parameters 258
Specific state variables, thermodynamics 87
Specifications, mechatronics 1644
Spectral luminous efficiency, lighting quantities 80
Spectral sensitivity, night vision systems 1877
Speed governing, Electronic Diesel Control 912
Speed limit, rolling bearings 379
Speed measurement, radar-sensor system 1819
Speed sensor for transmission control 1742
Speed symbol, tire designation 1182
Spherical lenses, geometric optics 76

Spherical snap-on connections 447
Spheroidization, heat treatment 335
Spinning-current principle, position sensors 1730
Spiral springs, springs 365
Spiral vee-form filter, gasoline filter 826
Spiral-type supercharger, mechanical superchargers 774
Spiral-wound filament, lighting equipment 1336
Spiral-wound mufflers, exhaust-gas system 795
Spitback test, evaporative-emissions test 1089
Splash oil, reciprocating-piston engine 701
Splitter box, step-up gear unit 624
Spoiler sunroof, sunroof systems 1406
Spool pressure-control valve, transmission control 638
Sport utility vehicles, overall motor-vehicle system 30
Sports Utility Cabriolet, vehicle bodies, passenger cars 1313
Sports Utility Vehicle, vehicle bodies, passenger cars 1313
Spot diameter, welding 450
Spot-welding electrodes, welding 451
Spray formation, high-pressure fuel injector 850
Spray formation, intake-manifold fuel injector 847
Spray formation, piezo high-pressure fuel injector 852
Spray propagation, diesel engine 658
Spray taper, intake-manifold fuel injector 847
Spray targeting, intake-manifold fuel injector 848
Spray targeting, intake-manifold injection 842
Spray targeting, piezo high-pressure fuel injector 852
Spray, intake-manifold fuel injector 847
Spray-directed combustion process, gasoline direct injection 845
Spreading-resistance principle 1763
Spring characteristic, springs 362
Spring combinations, springs 363
Spring damping, springs 362
Spring duty, springs 362
Spring element, snap-on connections 448
Spring ratio, motor-vehicle dynamics 489
Spring steel, metallic materials 276
Spring travel, snap-on connections 448
Spring-loaded brake system, brake systems, commercial vehicles 1232
Spring-mass system, acceleration sensor 1751
Spring-mounted point adjustment, suspension 1134
Springs with bellows, suspension 1130
Springs, machine parts 362
Spring-strut independent wheel suspension, wheel suspensions 1152
Spring-type brake cylinder, disk brake for commercial vehicles 1255
Sprockets, chain drives 734
Spur-gear differential, differentials 627
Spur-gear drives, step-up gear units 614
Spur-gear set, step-up gear units 610
Sputter bearing, friction bearings 373
Squeaking noises, vehicle acoustics 535
Squeeze-cast process 1164
Squish flows, diesel engine 658, 659
SRR, radar-sensor system 1822
Stabilization interventions, ESC 1277
Stabilizer spring, suspension 1130
Stabilizers, suspension 1131
Stages on the road to autonomous driving, future of automated driving 1936
Stainless steels, EN metallurgy standards 286
Stamped PowerFrame grid, starter batteries 1479
Standard architecture, driver-assistance systems 1807
Standard deviation, technical statistics 227
Standard drive, drivetrain with internal-combustion engine 598
Standard model library, mechatronics 1643
Standard potential, corrosion processes 340
Standardization of steels, EN metallurgy standards 286
Standardization of the iron casting materials, EN metallurgy standards 287
Standardization, CAN 1696
Standardization, MOST 1710
Standards for software in motor vehicles 1647
Standards, electromagnetic compatibility 1534

Standby heating,
preheating systems 955
Standing wave ratio,
electrical engineering 128
Star connection, alternators 1492
Star diagram,
reciprocating-piston engine 696
Star topology, vehicle communication networks 1689
Star washer, frictional joints 434
Star, vehicle communication networks 1689
Star-bus topology, vehicle communication networks 1690
Star-ring topology, vehicle communication networks 1690
Start heating, flame start systems 955
Start of Frame, CAN 1695
Start of injection,
axial-piston distributor pump 950
Start readiness,
start-assist systems 960
Start support, flame start system 960
Start/stop function, hybrid drives 992
Start/stop system, hybrid drives 994
Start/stop system,
starters for passenger cars 1513
Start-assist systems,
diesel engines 954
Starter batteries 1476
Starter battery,
12 V vehicle electrical system 1459
Starter generator, hybrid drives 995
Starter pinion,
starters for passenger cars 1509
Starters for heavy commercial vehicles,
automotive electrical systems 1514
Starters for passenger cars and light commercial vehicles,
automotive electrical systems 1506
Starting battery,
12 V vehicle electrical system 1459
Starting-torque compensation,
chassis systems 1118
Start-of-delivery control,
axial-piston distributor pump 950
Startup frame indicator, FlexRay 1699
Start-up process, FlexRay 1698
State diagram for carbon 192
State diagram for water 193
State diagrams, substances 192
State of charge,
12 V vehicle electrical system 1460
State of function,
12 V vehicle electrical system 1460
State of health,
12 V vehicle electrical system 1461
State variables, thermodynamics 86
States of aggregation, substances 192
Static cornering headlamps,
lighting equipment 1355
Static friction, mechanics 53
Static seals, seal technology 384
Static segment, FlexRay 1698
Static stress, threaded fasteners 442
Statics, mechanics 49
Stationary heat transmission,
thermodynamics 98
Statistical Process Control,
technical statistics 232
Stator, alternators 1492
Stator, hydrodynamic torque converter 603
Steady-state circular-course driving,
driving-dynamics test procedures 499
Steady-state voltage,
starter batteries 1488
Steel cord, radial tire 1172
Steel group number,
EN metallurgy standards 287
Steel springs, suspension 1125
Steel types, metallic materials 272
Steel wheel, wheels 1160
Steel wheels 1162
Steel-belt assembly, radial tire 1172
Steel-link chains, chain drives 730
Steels, metallic materials 271
Steering angle, chassis systems 1115
Steering axis, chassis systems 1118
Steering box, steering 1195
Steering column, steering 1195
Steering input, ESC 1275
Steering lock,
vehicle security systems 1444
Steering moment under braking conditions, Motorcycle Stability Control 1299
Steering parking assistant, parking and maneuvering systems 1884
Steering pinion, steering 1195
Steering ratio, chassis systems 1118
Steering, chassis systems 1192
Steering-angle sensor 1736
Steering-axis spread,
chassis systems 1118
Steering-wheel angle,
chassis systems 1118
Steering-wheel lock,
vehicle security systems 1444

Steering-wheel torque,
chassis systems 1118
Stefan-Boltzmann constant,
thermodynamics 100
Stefan-Boltzmann law,
thermodynamics 100
Step input, driving-dynamics test
procedures 500
Step-change sensors, λ sensors 1766
Step-down converter,
DC-DC converters 1595
Step-index fiber,
optical fibers/waveguides 83
Stepped recess,
reciprocating-piston engine 676
Step-up converter,
DC-DC converters 1595
Step-up gear unit, drivetrain with
internal-combustion engine 597
Step-up gear units, drivetrain 608
Steradian, lighting quantities 82
Steradian, SI units 34
Stick-slip behavior, lubricants 539
Stick-slip effect, friction bearings 370
Stiffness optimization,
finite-element method 240
Stimulated emission,
laser technology 83
Stochastic course of irregularity,
motor-vehicle dynamics 487
Stochastic-graphical models,
sensor-data fusion 1869
Stoichiometric ratio, mixture formation,
spark-ignition engines 838
Stop lamps, lighting equipment 1361
Stoppie,
Motorcycle Stability Control 1298
Stopping time,
motor-vehicle dynamics 476
Stop-sign assistant,
intersection assistant 1927
Storage in compressed form,
engines fueled by natural gas 894
Storage of LPG 887
Storage of natural gas,
engines fueled by natural gas 894
Storage technologies, batteries
for electric and hybrid drives 1607
Straight line, mathematics 205
Straight toothing, step-up gear units 609
Strain gage, pressure sensors 1757
Strain-gage resistor,
pressure sensors 1757
Strain-gage technology,
pressure sensors 1757
Stratified-charge cloud,
gasoline direct injection 844
Stratified-charge operation,
gasoline direct injection 844
Strength members, radial tire 1171
Strength verification, mechanics 51
Stress corrosion cracking,
types of corrosion 341
Stress value, elastomer materials 421
Stress-free annealing,
heat treatment 335
Stress-strain curve,
material parameters 260
Stretch Fit drives, belt drives 725
Stribeck curve 699
Stribeck curve, friction bearings 371
Stribeck curve, lubricants 543
Stripline,
electromagnetic compatibility 1531
Stroke/connecting-rod ratio,
reciprocating-piston engine 692
Strong electrolyte,
reactions of substances 198
Structural compositions of steels,
metallic materials 271
Structural design,
corrosion protection 344
Structural steel, metallic materials 272
Structure modification,
modal analysis 65
Structure switchovers,
control engineering 245
Structure-borne sound vibrations,
knocking 1755
Styrene butadiene rubber,
elastomer materials 416
Sublimation curve, water 193
Substance concentrations,
chemistry 194
Substance term, chemistry 192
Substances, chemistry 192
Substitute functions,
Electronic Diesel Control 914
Substitution, mathematics 217
Substrate coating,
catalytic converter 874
Subsystems,
Motronic system overview 802
Subtracting amplifier,
operational amplifier 153
Sub-transmissions,
step-up gear unit 624
Suction jet pump,
fuel-supply module 829
Sulfating, starter batteries 1484

Sulfur content, fuels 562
Sulfur oxides,
exhaust-gas emissions 717
Sun gear, step-up gear units 610
Sunroof glazing made from
laminated safety glass,
automotive glazing 1377
Sunroof systems 1406
Super fuels, gasoline 563
Supercapacitors, electromobility 1616
Superchargers (mechanically
driven), turbochargers and
superchargers 772
Supercharging processes 671
Supercharging ratio 671
Supercharging 664, 671
Supercritical phase, water 193
Superimposed angle, steering 1202
Superimposed steering, steering 1200
Superposition principle, wave optics 78
Supervised learning,
artificial intelligence 251
Supervised training,
sensor-data fusion 1868
Supplementary A/C units 1394
Supplementary diodes,
alternators 1493
Supplementary functions, automatic
brake-system operations 1292
Supplementary transverse-dynamics
functions, ESC 1280
Supply coupling head, brake systems,
commercial vehicles 1230
Supply volume, brake systems,
commercial vehicles 1233
Support ring, O-ring 388, 392
Supporting interventions,
lane assistance 1908
Surface corrosion,
types of corrosion 342
Surface deviations, tolerances 354
Surface engineering,
deposits and coatings 348
Surface filters, air-intake system 756
Surface hardening, heat treatment 333
Surface insulators, plastics 323
Surface micromechanics 1721
Surface parameters, tolerances 357
Surface pressure, frictional joints 433
Surface pressure, mechanics 51
Surface pressure,
threaded fasteners 442
Surface roughness 701
Surface roughness,
radial shaft seal 401
Surface-air-gap spark concept,
spark plug 870
Surface-gap spark concept,
spark plug 870
Surface-micromechanical acceleration
sensors 1754
Surface-micromechanical yaw-rate
sensors 1744
Surge damping control,
Electronic Diesel Control 913
Surge limit,
exhaust-gas turbocharger 778
Survival probability,
technical statistics 229
Susceptibility, magnetic materials 290
Suspension design elements,
suspension 1124
Suspension system, suspension 1124
Suspension systems, suspension 1132
Suspension, chassis systems 1124
Suspension, substances 192
Sustaining test,
energy demand for drives 509
SUVs, overall motor-vehicle system 30
Sweating, soldering 454
Swelling behavior,
elastomer materials 415
Swept volume,
reciprocating-piston engine 706
Swirl duct, air-intake system 765
Swirl pot, fuel-supply module 829
Swirl, air-intake system 766
Swirl, internal-combustion
engines 653, 659
Switch, Ethernet 1704
Switchable glazing,
automotive glazing 1378
Switching diode, electronics 141
Switching frequency,
DC-DC converters 1599
Switching relay, starters for heavy
commercial vehicles 1516
Switching valve,
transmission control 635
Switch-off function,
starters for passenger cars 1511
Switch-on conditions,
on-board diagnostics 1097
Symbol Window, FlexRay 1698
Sync frame indicator, FlexRay 1699
Synch Break, LIN 1701
Synch Field, LIN 1701
Synchromesh, step-up gear units 611
Synchronization break, LIN 1701
Synchronization field, LIN 1701

Synchronization, FlexRay 1698
Synchronization, LIN 1702
Synchronizer ring,
step-up gear units 611
Synthetic fuels, fuels 583
Synthetic oil, lubricants 543
System architecture, automotive
software engineering 1654
System architecture, integrated driving-
dynamics control systems 1307
System Control, Motronic 804
System Document, Motronic 804
System lock, locking systems 1439
System of coordinates,
chassis systems 1113
System of coordinates,
vehicle motion 462
system of units, quantities and units 32
Systems of linear equations,
Mathematics 224
Systems, vehicle electrical systems
for commercial vehicles 1469

T

Tail lamp,
lighting equipment 1340, 1360
Tailored blanks, wheels 1160
Tandem brake master cylinder,
passenger-car brake system 1220
Tandem fuel pump,
low-pressure diesel fuel supply 919
Tandem-axle assemblies, vehicle
bodies, commercial vehicles 1326
Tangent function, mathematics 209
Tangential acceleration,
vehicle motion 463
Tangential sensors, rpm sensors 1741
Tank ventilation, fuel supply 824
Tank-breathing losses,
evaporative-emissions test 1088
Tank-leak diagnosis,
on-board diagnostics 1099
Tank-to-Wheel efficiency,
energy demand for drives 514
Tank-to-Wheel path,
energy demand for drives 514
Tank-ventilation gases,
air-intake system 766
Tapered connection,
frictional joints 433
Tapered rolling bearing,
rolling bearings 378
Tapered spray,
intake-manifold fuel injector 847
Taper-lock joints, frictional joints 435
Taper-lock ring, frictional joints 434
Taut strand, mechanics 54
TCS control,
wheel-slip control systems 1264
TCS engine-control intervention
in commercial vehicles,
traction control system 1271
TCS engine-control intervention
in passenger cars,
traction control system 1268
TCS systems for commercial vehicles,
wheel-slip control systems 1269
TCS systems for passenger cars,
wheel-slip control systems 1265
TCS valve, traction control system 1271
Tear-propagation strength,
elastomer materials 421
Technical obstacles,
future of automated driving 1935
Technical statistics,
mathematics and methods 226
Technological active chain, architecture
of electronic systems 1676
Technology model, architecture
of electronic systems 1676
Telegrams, vehicle communication
networks 1682
Telescopic shock absorbers,
shock absorbers 1136
TEM cell,
electromagnetic compatibility 1531
TEM waveguides,
electromagnetic compatibility 1531
Temperature blending flaps,
climate control 1393
Temperature coefficient of coercive field
strength, material parameters 259
Temperature coefficient of magnetic
polarization, material parameters 259
Temperature coefficient 1762
Temperature coefficient,
electrical engineering 107
Temperature curve,
temperature sensor 1761
Temperature ranges in motor
vehicles 1760
Temperature sensors 1760, 1763
Temperature, legal units 38
Tempering steel, metallic materials 272
Tensile strength,
elastomer materials 420
Tensile strength, locking systems 1434
Tensile strength,
material parameters 260

Tensile strength, mechanics 51
Tensile stress, mechanics 52
Tensile stress-strain properties, elastomer materials 420
Tensile value, locking systems 1434
Tension member, belt drives 727
Tension members, belt drives 721
Tension pulleys, belt drives 724
Tension spring, radial shaft seal 399
Tensioning rail, chain drives 735
Terminal designations 1557
Terminal diagram, circuit diagrams 1550
Terminal stud, spark plug 866
Terminal voltage, starter batteries 1483
Terminating resistor, electrical engineering 128
Tesla, SI units 34
Test and hedging, driver-assistance systems 1809
Test bench, exhaust-gas measuring techniques 1079
Test cycles for heavy commercial vehicles, emission-control legislation 1067
Test cycles for passenger cars and light commercial vehicles, emission-control legislation 1057
Test of software and ECUs, automotive software engineering 1660
Test procedures, emission-control legislation 1042
Test-bench measuring techniques, exhaust-gas measuring techniques 1082
Tester-based diagnostic modules, ECU diagnostics 1106
Tetrahedrons, finite-element method 236
Theft deterrence, locking systems 1438
Theft-alarm system, vehicle security systems 1448
Theoretical spring length, suspension 1128
Thermal acceleration sensors 1752
Thermal aging process, vehicle electrical systems for hybrid and electric vehicles 1564
Thermal aging, catalytic converter 876
Thermal conduction, thermodynamics 96
Thermal conductivity, material parameters 258
Thermal conductivity, thermodynamics 97
Thermal control, air-intake system 759
Thermal engines, internal-combustion engines 642
Thermal equation of state, thermodynamics 89
Thermal ignition, spark plug 868
Thermal layout, electric final-drive unit 1021
Thermal limitation, electrical machines 1581
Thermal management for electric drives, electrification of drive 980
Thermal Management subsystem, control of electric drives 988
Thermal management, batteries for electric and hybrid drives 1613
Thermal management, engine cooling 749
Thermal management, engine lubrication 753
Thermal management, fuel cell 1036
Thermal radiation, thermodynamics 96, 100
Thermal runaway, batteries for electric and hybrid drives 1611
Thermal-conduction resistance, thermodynamics 97
Thermochemical treatment, heat treatment 337
Thermocouples, electrical engineering 131
Thermodynamic systems 86
Thermodynamics 86
Thermoelectric method, hydrogen sensors 1775
Thermoelectric series, electrical engineering 131
Thermoelectricity, electrical engineering 131
Thermoelectromotive force, electrical engineering 132
Thermomagnetic effects, electrical engineering 133
Thermoplastic elastomers, plastics 315
Thermoplastic plastics, chemical name 308
Thermoplastics, mechanical properties 312
Thermoplastics, plastics 307
Thermosets, plastics 319
Thermostatic switch, starters for heavy commercial vehicles 1517
Thick-film resistors, temperature sensors 1762

Thin-film metallic resistors, temperature sensors 1761
Third law, thermodynamics 89
Third-octave band spectrum, acoustics 68
Third-order reaction, reaction kinetics 197
Thixotropy, lubricants 542
Thorax bag, occupant-protection systems 1421
Thread dimensions, threaded fasteners 445
Threaded fasteners 437
Threaded-fastener locking devices, threaded fasteners 443
Three-box design, vehicle bodies, passenger cars 1316
Three-cell limit-current sensor, NO_x sensor 1772
Three-phase alternating current, alternators 1492
Three-phase alternating-current rectification, alternators 1493
Three-phase bridge connection, inverter 1583
Three-phase chargers, on-board charger 1605
Three-phase current system 178
Three-phase current, alternators 1492
Three-phase system, three-phase current system 178
Three-phase voltage system, inverters 1587
Three-plunger radial-piston pump, common rail 937
Three-point belt, occupant-protection systems 1418
Three-sensor λ control, exhaust-gas treatment 880
Three-way catalytic converter, catalytic exhaust-gas treatment 874
Three-zone climate control 1393
Threshold-value controllers, shock absorbers 1142
Throttle device, cylinder charge 813
Throttle map, cylinder charge 813
Throttle valve, cylinder charge 813
Throttle-valve angle sensor 1725
Throughflow characteristics, shock absorbers 1137
Tie rod, steering 1199
Tier 2, EPA legislation 1048
Tier 3, EPA legislation 1048
Tightening torques, threaded fasteners 441
Tilt sensor, theft-alarm system 1449
Time control, FlexRay 1698
Time control, vehicle communication networks 1686
Time master, vehicle communication networks 1686
Time quantities, legal units 36
Time slot, FlexRay 1698
Time to brake, emergency-braking systems 1917
Time to brake, lane assistance 1911
Time to collision, emergency-braking systems 1917
Time to steer, emergency-braking systems 1917
Time-controlled single-cylinder pump injection systems 944
Time-dependent current, electrical engineering 114
Time-of-flight principle, lidar-sensor system 1828
Time-triggered CAN 1696
Timing chains, chain drives 730
Timing device, axial-piston distributor pump 949
Timing devices, radial-piston distributor injection pump 952
Timing master, MOST 1708
Timing, charge cycle 665
Timing, cylinder charge 814
Timing-chain drive, reciprocating-piston engine 687
Tin bronze, nonferrous metals 277
Tin deposits, deposits and coatings 349
Tinted glazing, automotive glazing 1376
Tipping resistance, driving dynamics – commercial vehicles 494
Tire aligning force, chassis systems 1121
Tire caster offset, chassis systems 1121
Tire construction 1171
Tire designation 1182
Tire flank, radial tire 1171
Tire grip 1177
Tire inflation pressure 1173
Tire label 1185
Tire noise, tires 1186
Tire tests 1187
Tire tread, radial tire 1174
Tire-pressure control, tires 1190
Tire-pressure monitoring system, architecture of electronic systems 1667

Tire-pressure monitoring systems 1188
Tires 1170
Tires, suspension 1124
Titanium alloys,
nonferrous metals 280
Titanium aluminum nitride,
deposits and coatings 352
Titanium carbonitride,
deposits and coatings 352
Titanium nitride,
deposits and coatings 352
Titanium, nonferrous metals 280
Toe angle, chassis systems 1114
Toe-in, chassis systems 1115
Toe-out, chassis systems 1115
Token passing, vehicle
communication networks 1687
Tolerance class, tolerances 355
Tolerance ring, frictional joints 434
Tolerances, machine parts 354
Tolerancing principles, tolerances 354
Tonal quality, acoustics 70
Tool steel, metallic materials 272
Tool techniques,
two-component seals 409
Toothed belt, belt drives 726
Toothed wheels, belt drives 727
Toothed-belt drive,
reciprocating-piston engine 687
Toothed-belt profiles, belt drives 727
Tooth-type chain, chain drives 731
Top dead center, charge cycle 664
Top-down approach, architecture
of electronic systems 1679
Top-feed principle, injector 891
Topologies, FlexRay 1697
Topologies, hybrid drives 998
Topology model, architecture
of electronic systems 1677
Topology, Ethernet 1704
Topology, MOST 1709
Topology, Serial Wire Ring 1711
Torque blending,
electrification of drive 977
Torque coordinator,
Electronic Diesel Control 910
Torque curve, supercharging 672
Torque Demand subsystem,
control of electric drives 985
Torque Demand, Motronic 802
Torque distribution in the longitudinal
direction in four-wheel drives 1305
Torque measurement, sensors 1764
Torque path,
Electronic Diesel Control 910
Torque position,
reciprocating-piston engine 703
Torque rise,
reciprocating-piston engine 705
Torque sensor 1764
Torque sensor, steering 1200
Torque Structure subsystem,
control of electric drives 987
Torque structure,
Electronic Diesel Control 910
Torque Structure, Motronic 802
Torque, mechanics 47
Torque-conversion factor,
hydrodynamic torque converter 604
Torque-vectoring actuator 1306
Torsion damper, vibration isolation 606
Torsion snap-on hooks,
snap-on connections 447
Torsion springs, springs 365
Torsional force,
reciprocating-piston engine 695
Torsional stress, mechanics 52
Torsional stresses, springs 367
Torsional vibrations,
reciprocating-piston engine 689
Torsional-link axles,
wheel suspensions 1149
Torsion-bar springs, springs 366
Torsion-bar springs, suspension 1127
Total braking time 1210
Total contamination, diesel fuel 573
Total failure rate, technical statistics 229
Total reflection, geometric optics 76
Total running resistance,
vehicle physics 464
Total-loss lubrication,
reciprocating-piston engine 688
Touchscreen, display and control 1792
Tour buses (coaches), vehicle bodies,
commercial vehicles 1330
Tox clinching,
penetration-clinching processes 459
TPEG standard, traffic telematics 1801
Track width, chassis systems 1114
Tracking control, ACC 1892
Tracking,
sensor-data fusion 1861, 1868
Traction battery, batteries for electric
and hybrid drives 1606
Traction control system, wheel-slip
control systems 1261, 1264
Traction-electrical-system
DC-DC converters 1597
Tractive force, drivetrain with
internal-combustion engine 596

Tractive resistance equation, aerodynamics 519
Tractive-force curves, drivetrain with internal-combustion engine 595
Tractive-force hyperbola, drivetrain with internal-combustion engine 595
Tractive-force running diagram, motor-vehicle dynamics 472
Tractive-force-interrupted transmissions, step-up gear units 615
Tractive-force-interruption-free gear change, manually shifted transmissions 618
Tractor units, vehicle bodies, commercial vehicles 1325
Trade-off, automotive software engineering 1662
Traffic Message Channel, traffic telematics 1801
Traffic Message Channel, vehicle navigation 1875
Traffic messages, traffic telematics 1802
Traffic telematics 1801
Traffic-jam assistant, future of automated driving 1936
Traffic-jam pilot, future of automated driving 1937
Traffic-light assistant, intersection assistant 1927
Trailer control module, electronically controlled brake system 1245
Trailer control valve, brake systems, commercial vehicles 1241
Trailer operation, vehicle bodies, commercial vehicles 1331
Trailer sway mitigation, ESC 1281
Trailer Sway Mitigation, ESC 1282
Trailer, FlexRay 1700
Trailing axles, vehicle bodies, commercial vehicles 1326
Trailing link, wheel suspensions 1148
Trailing-link independent wheel suspension, wheel suspensions 1150
Transaxle drive, drivetrain with internal-combustion engine 599
Transceiver, vehicle communication networks 1684
Transceiver, vehicle immobilizer 1445
Transfer elements, control engineering 243
Transfer function, vibrations and oscillations 61
Transfer passage, 2-stroke process 666
Transfer rate, LIN 1701
Transfer rate, MOST 1707
Transfer rates, Ethernet 1703
Transfer reliability, vehicle communication networks 1691
Transfer system, Ethernet 1703
Transfer system, LIN 1701
Transfer system, MOST 1707
Transfer system, PSI5 1705
Transfer technique, two-component seals 410
Transfer time, motor-vehicle dynamics 475
Transfer-oriented method, vehicle communication networks 1688
Transformation temperature, glass 303
Transgranular corrosion, types of corrosion 341
Transient response, driving-dynamics test procedures 500
Transient voltage, ignition coil 863
Transistors, electronics 144
Transition fit, rolling bearings 380
Transition range, motor-vehicle dynamics 480
Transition test, driving-dynamics test procedures 502
Translational motions, vehicle motion 462
Transmission acoustics 533
Transmission control, drivetrain 630
Transmission device, brake systems 1207
Transmission device, passenger-car brake system 1221
Transmission media, FlexRay 1697
Transmission of force, mechanics 50
Transmission oils, lubricants 548
Transmission paths, traffic telematics 1801
Transmission ratio, high-pressure pump, common rail 938
Transmission Start Sequence, FlexRay 1700
Transmitter initiative, CAN 1695
Transponder, vehicle immobilizer 1445
Transport layer, vehicle communication networks 1685
Transport mechanisms, fuel cell 1032
Transport Protocol Experts Group, traffic telematics 1801
Transversal waves, wave optics 78
Transverse pole, chassis systems 1115
Transverse press fit, frictional joints 432
Transverse tilting moment, reciprocating-piston engine 695

Transverse-dynamics control, ESC 1277
Transverse-dynamics controller, ESC 1275
Transverse-pin joint, positive or form-closed joints 429
Trapezoidal-link independent wheel suspension, wheel suspensions 1152
Travel-direction determination, vehicle navigation 1873
Traversing cradle, wind tunnels 527
Tread grooves, tire tread 1174
Tread pattern, radial tire 1172
Tread wear indicator, tire tread 1174
Trenching, micromechanics 1721
Triangle of forces, mechanics 49
Tribological system, reciprocating-piston engine 699
Tribology, reciprocating-piston engine 699
Triggering, starters for passenger cars 1512
Trigonometric functions, mathematics 207
Trilateration, GPS 1872
Trimming resistor, temperature sensor 1761
Trimming, temperature sensor 1761
Trimolecular reaction, reaction kinetics 197
Triple bonds, chemical bonds 189
Triple point, water 193
Triple-barrel radial-piston pump, gasoline direct injection 832
Triple-cone synchromesh, step-up gear units 612
Triple-shaft transmissions, step-up gear units 616
Trough flap, air-intake system 765
Truck chassis, vehicle bodies, commercial vehicles 1325
Truck electrical system, vehicle electrical systems for commercial vehicles 1466
TRX rim, wheels 1166
TSG panes, automotive glazing 1375
T-shaped topology, electric final-drive unit 1023
Tube/fin systems, water cooling 738
Tubular body frame, finite-element method 239
Tubular heating element, preheating systems 955
Tubular rivet, riveting 458
Tumble flap, air-intake system 764
Tumble, air-intake system 765
Tumble, gasoline engine 653
Tuned-intake-tube charging, air-intake system 764
Tungsten inert-gas welding, welding 452
Tungsten, lighting equipment 1336
Tunnel magnetoresistive effect, electrical engineering 137
Turbine, exhaust-gas turbocharger 779
Turbine, hydrodynamic torque converter 603
Turbocharger, exhaust-gas turbocharging 672
Turbochargers and superchargers for internal-combustion engines 772
Turbulence flap, air-intake system 764
Turbulence intensity, gasoline engine 651
Turbulences, diesel engine 657
Turbulences, gasoline engine 651
Turbulent flow, flow mechanics 57
Turbulent kinetic energy, diesel engine 659
Turbulent kinetic energy, gasoline engine 652
Turcon PTFE rotary seal, tire-pressure control 1190
Turcon Roto L, tire-pressure control 1190
Turn collision warning for trucks, information and warning systems 1901
Turning circle, chassis systems 1120
Turning tool, two-component seals 409
Turning-circle regulation, driving dynamics - commercial vehicles 495
Turn-signal lamps, lighting equipment 1359
Turnstile, two-component seals 410
Twin cab, vehicle bodies, commercial vehicles 1324
Twin rocker-arm assembly, reciprocating-piston engine 685
Twin-tube shock absorbers, shock absorbers 1137
Twist-beam axles, wheel suspensions 1149
Twisted Nematic LCD, display and control 1789
Twisted pair, buses in motor vehicles 1692
Two-actuator system, common rail 927
Two-battery vehicle electrical system, 12 V vehicle electrical system 1459

Two-box design, vehicle bodies, passenger cars 1316
Two-component adhesives, adhesive technologies 455
Two-component seals, seals 408
Two-detent capability, locking systems 1432
Two-layer piezoceramics, acceleration sensor 1753
Two-plate model, lubricants 541
Two-plunger in-line piston pump, common rail 942
Two-plunger radial-piston pump, common rail 939
Two-point links, wheel suspensions 1152
Two-point λ sensor, exhaust-gas treatment 879
Two-point λ sensors 1767
Two-sensor λ control, exhaust-gas treatment 880
Two-stage exhaust-gas cooling 745
Two-stage meshing, starters for heavy commercial vehicles 1515
Two-step λ control, exhaust-gas treatment 879
Two-stroke combustion processes, charge cycle 666
Two-stroke cycle, charge cycle 666
Two-track model, motor-vehicle dynamics 483
Two-wire cable, CAN 1693
Two-zone temperature control, climate control 1393
Type approval, electromagnetic compatibility 1534
Type approval, emission-control legislation 1042
Type approval, EU legislation 1052
Type designation, spark plug 871
Type of characteristic curve, sensors 1718
Type tests, EU legislation 1053
Types of cast iron, metallic materials 277
Types of corrosion, corrosion and corrosion protection 341
Types of spring, suspension 1127
Types of support, lane assistance 1909
Types of wind tunnel 524

U

Ultra-close range, driver-assistance systems 1809
Ultra-long-range lidar, lidar-sensor system 1827
Ultrasonic parking aid, parking and maneuvering systems 1882
Ultrasonic parking assistant, parking and maneuvering systems 1884
Ultrasonic sensor system, driver assistance and automated driving 1812
Ultrasonic sensor, ultrasonic sensor system 1812
Ultrasound, acoustics 66
Unalloyed high-grade steels, EN metallurgy standards 286
Unalloyed stainless steels, EN metallurgy standards 286
Unalloyed steels, EN metallurgy standards 286
Uncontrolled combustion, gasoline engine 653
Undamped natural angular frequency, motor-vehicle dynamics 488
Underexcitation, synchronous machine 171
Underfloor catalytic converter, catalytic exhaust-gas treatment 877
Underfloor-engine drive, drivetrain with internal-combustion engine 599
Understeering, motor-vehicle dynamics 482
Uniball joints, chassis systems 1124
Uniform strain, material parameters 261
Uniform-distribution requirement, air-intake system 766
Unit bore, tolerances 355
Unit injector, commercial vehicles 946
Unit injector, passenger cars 944
Unit matrix, mathematics 224
Unit pump, commercial vehicles 947
Unit shaft, tolerances 355
Unit-injector system 944
Unit-injector system, commercial vehicles 946
Unit-injector system, passenger cars 944
Unitized body, vehicle bodies, commercial vehicles 1331
Unit-pump system 947
Unit-pump system, commercial vehicles 947
Units, quantities and units 32

Unlocking zone, locking systems 1441
Unscrewing, threaded fasteners 443
Unsupervised learning,
artificial intelligence 251
Upshift curve, transmission control 633
Urea/water solution, exhaust-gas
treatment, diesel engines 968
Urea/water solution,
operating fluids 590
US legislation, emission legislation
for motorcycles 1071
US legislation, emission-control
legislation for heavy commercial
vehicles 1062
US sealed beam,
lighting equipment 1340
US test cycles,
emission-control legislation 1057
US06 cycle, test cycles 1058
Useful speed range,
reciprocating-piston engine 704
User addressing, vehicle
communication networks 1688
User data, vehicle
communication networks 1688
User-oriented method, vehicle
communication networks 1688
Using the battery, starter batteries 1487
Utility factor,
energy demand for drives 509
UV/VIS spectroscopy,
chemical bonds 191

V

V model, automotive software
engineering 1650
V model, mechatronics 1643
Vacuum brake booster,
passenger-car brake system 1217
Vacuum carburizing,
heat treatment 337
Vacuum non-return valve,
passenger-car brake system 1218
Vacuum units, air-intake system 766
Vacuum-operated boosters,
passenger-car brake system 1217
Valence electrons, periodic table 182
Valet parking,
future of automated driving 1937
Validation, mechatronics 1644
Valve acceleration,
reciprocating-piston engine 686
Valve arrangements,
reciprocating-piston engine 684, 686
Valve gear,
reciprocating-piston engine 685
Valve guide,
reciprocating-piston engine 686
Valve overlap period, charge cycle 665
Valve overlap, cylinder charge 814
Valve overlap, variable valve timing 667
Valve seat,
reciprocating-piston engine 686
Valve spring,
reciprocating-piston engine 685
Valve timing, charge cycle 665
Valve timing, cylinder charge 814
Valve timing,
reciprocating-piston engine 686
Valve velocity,
reciprocating-piston engine 686
Valve-lift curve,
reciprocating-piston engine 686
Valve-rotating system,
reciprocating-piston engine 686
Valves, reciprocating-piston engine 678
Valve-timing designs,
reciprocating-piston engine 685
Valve-timing diagram,
reciprocating-piston engine 686
Van der Waals equation,
thermodynamics 90
Van der Waals forces,
chemical bonds 190
Vane pump,
low-pressure diesel fuel supply 918
Vapor lock, brake fluids 556
Vapor lock, gasoline 565
Vapor pressure, gasoline 566
Vapor/liquid ratio, gasoline 567
Vaporization curve, water 193
Vapor-lock index, gasoline 567
Vapor-phase inhibitors,
corrosion protection 346
Variable compression ratio,
reciprocating-piston engine 682
Variable induction system,
air-intake system 763
Variable turbine geometry,
exhaust-gas turbocharger 782
Variable turbine geometry,
supercharging 672
Variable valve timing 667
Variance, technical statistics 227
Variator, step-up gear unit 622
V-belts, belt drives 720
Vectors, mathematics 213
Vegetable oils, diesel fuels 576
Vehicle access, locking systems 1441

Vehicle acoustics 530
Vehicle aerodynamics 516
Vehicle aerodynamics, vehicle bodies, passenger cars 1317
Vehicle architecture, vehicle bodies, passenger cars 1312
Vehicle bodies, commercial vehicles 1324
Vehicle bodies, passenger cars, vehicle bodies 1310
Vehicle body shapes, vehicle bodies, passenger cars 1314
Vehicle center plate, vehicle motion 462
Vehicle classes, CARB legislation 1043
Vehicle classes, emission-control legislation for heavy commercial vehicles 1062, 1064, 1066
Vehicle classes, EPA legislation 1047
Vehicle classes, EU legislation 1050
Vehicle classes, Japanese legislation 1055
Vehicle communication networks, automotive electronics 1682
Vehicle concept, vehicle bodies, passenger cars 1312
Vehicle control unit for electric vehicles, control of electric drives 984
Vehicle control unit, control of a hybrid drive 1010
Vehicle Distributed Executive 1648
Vehicle Dynamics Management 1282, 1304
Vehicle electrical systems for commercial vehicles, automotive electrical systems 1466
Vehicle electrical systems for hybrid and electric vehicles, electrics for electric and hybrid drives 1560
Vehicle Functions subsystem, control of electric drives 988
Vehicle handling, driving-dynamics test procedures 498
Vehicle headlamp aiming device, lighting equipment 1345
Vehicle integration, electric final-drive unit 1024
Vehicle Motion Control, architecture of electronic systems 1673
Vehicle navigation 1872
Vehicle nominal motion, ESC 1277
Vehicle package, vehicle bodies, passenger cars 1313
Vehicle power supply control unit, 12 V vehicle electrical system 1459
Vehicle recall, on-board diagnostics 1098
Vehicle response, chassis systems 1110
Vehicle safety, safety systems in the motor vehicle 1410
Vehicle segments, vehicle bodies, passenger cars 1313
Vehicle Stability Assist, driving-dynamics control system 1287
Vehicle Stability Control, driving-dynamics control system 1287
Vehicle stability, ESC 1273
Vehicle systems, vehicle bodies, passenger cars 1315
Vehicle testing procedures, electromagnetic compatibility 1532
Vehicle to Grid, on-board charger 1602
Vehicle to Home, on-board charger 1602
Vehicle to Load, on-board charger 1602
Vehicle vibration characteristics, chassis systems 1112
Vehicle wind tunnels 524
Vehicle-centralized E/E architecture, architecture of electronic systems 1669
Vehicle-concept development, vehicle bodies, passenger cars 1317
Vehicle-electrical-system simulation, 12 V vehicle electrical system 1458
Vehicle-speed controller 1890
Vehicle-vehicle communication, traffic telematics 1803
Velocity coefficient, flow mechanics 58
Velocity of sound, acoustics 67
Velocity, mechanics 46
V-engine, reciprocating-piston engine 690
Vent flaps, climate control 1393
Ventilation openings, starter batteries 1478
Ventilation vent, climate control 1393
Ventilation, aerodynamics 521
Venturi mixer, LPG operation 887
Versions of Motronic 805
Vertical shaft, reciprocating-piston engine 687
Very low-emission vehicles, emission-control legislation for heavy commercial vehicles 1066
VGA resolution, night vision systems 1876
Vibration absorbers, chassis systems 1144

Vibration absorbers,
exhaust-gas system 796
Vibration absorption 64
Vibration corrosion cracking,
types of corrosion 341
Vibration damper,
chassis systems 1136
Vibration damper,
reciprocating-piston engine 689
Vibration damping 63
Vibration isolation 63
Vibration isolation, drivetrain 605
Vibration minimization,
chassis systems 1111
Vibration reduction 63
Vibration resistance,
material parameters 261
Vibration sensors 1750
Vibrational stress,
threaded fasteners 442
Vibration-proof battery,
starter batteries 1486
Vibrations and oscillations 60
Vickers hardness, heat treatment 330
Video systems, parking and
maneuvering systems 1886
Virtual function bus system, automotive
software engineering 1648
Virtual prototyping,
vehicle acoustics 536
Virtual reality, vehicle bodies,
passenger cars 1323
Virtual safety shield,
driver-assistance systems 1809
Viscoelasticity, tire grip 1177
Viscosimetric quantities, legal units 38
Viscosity at very low temperatures,
brake fluids 557
Viscosity grades, lubricants 542, 543
Viscosity index, lubricants 543
Viscosity, brake fluids 557
Viscosity, diesel fuel 572
Viscosity, flow mechanics 57
Viscosity, lubricants 543
Viscosity,
shock-absorber systems 1139
Viscous clutch, water cooling 740
Visibility range,
motor-vehicle dynamics 478
Visible radiation, lighting quantities 80
Visual channel,
display and control 1784
Visual range,
lighting equipment 1345, 1356
V-joint brazing, soldering 455
Voice control, display and control 1793
Volatile corrosion inhibitors,
corrosion protection 346
Volatility, gasoline 565
Volt, SI units 34
Voltage divider,
temperature sensor 1761
Voltage law, electrical engineering 110
Voltage level, CAN 1693
Voltage regulation,
12 V vehicle electrical system 1455
Voltage regulation, alternators 1496
Voltage regulator,
electronic control unit (ECU) 1630
Voltage supply, ignition coil 862
Voltage transformer,
DC-DC converters 1594
Voltage, electrical engineering 105, 117
Volume blending,
electrification of drive 977
Volume change,
elastomer materials 415
Volume elements,
finite-element method 236
Volume lock, locking systems 1434
Volume micromechanics 1720
Volume, legal units 35
Volume-change work,
thermodynamics 87
Volumetric efficiency, charge cycle 664
Volumetric efficiency, high-pressure
pump, gasoline direct injection 831
Volumetric-flow map,
supercharging 673
VR-engine,
reciprocating-piston engine 690
Vulcanization accelerators,
radial tire 1171
Vulcanization, plastics 316

W

W recess,
reciprocating-piston engine 676
Wafer bonding, micromechanics 1722
Wake turbulence, aerodynamics 523
Wake-up, FlexRay 1700
Wake-up, LIN 1703
Wall film, intake-manifold injection 842
Wall-directed combustion process,
gasoline direct injection 845
Warm-up cycles,
on-board diagnostics 1097
Warning elements, parking and
maneuvering systems 1883

Warning flashers,
lighting equipment 1359
Warning lamp, diagnostics 1097
Washcoat, catalytic converter 874
Water consumption,
starter batteries 1483
Water cooling system,
reciprocating-piston engine 689
Water cooling, engine cooling 737
Water decomposition,
starter batteries 1483
Water pump,
reciprocating-piston engine 678
Water separation,
air-intake system 758, 769
Water separation, diesel filter 923
Water, Refrigerants 592
Water/glycol mixture, coolants 554
Water-absorption capacity,
brake fluids 556
Water-repellent glazing,
automotive glazing 1378
Watt linkage, wheel suspensions 1148
Watt, SI units 34
Watt-second, legal units 37
Wave optics 77
Wave propagation,
electrical engineering 129
Wave, vibrations and oscillations 61
Waveguide, electrical engineering 128
Wavelength 67
Wear marker, tire tread 1174
Wear resistance,
reciprocating-piston engine 699
Wear-detection aid, tire tread 1174
Weave test, driving-dynamics
test procedures 502
Weber, SI units 34
Wedge angle, mechanics 50
Wedge principle, mechanics 50
Wedge, mechanics 50
Wedge-actuated brake,
drum brake 1258
Weibull distribution,
technical statistics 229
Weight optimization,
finite-element method 240
Welded rail, common rail 943
Welding, permanent joints 450
Well-to-Tank efficiency,
energy demand for drives 512
Well-to-Tank path,
energy demand for drives 512
Well-to-Wheel analysis,
energy demand for drives 512
W-engine,
reciprocating-piston engine 690
Wet boiling point, brake fluids 557
Wet braking, tires 1175
Wet clutch,
power take-up elements 601
Wet grip, tire label 1186
Wet liner,
reciprocating-piston engine 682
Wet-sump lubrication,
reciprocating-piston engine 688
Wheatstone bridge,
pressure sensors 1757
Wheel brakes, chassis systems 1250
Wheel contact point,
chassis systems 1113
Wheel control, ESC 1277
Wheel controller, ESC 1275
Wheel design variations, wheels 1165
Wheel designation, wheels 1159
Wheel disk, wheels 1154
Wheel load, chassis systems 1114
Wheel mounting, wheels 1166
Wheel nave, wheels 1154
Wheel spider, wheels 1165
Wheel suspensions,
chassis systems 1146
Wheel trims, wheels 1166
Wheelbase, chassis systems 1114
Wheel-pressure modulator,
electrohydraulic brake 1225
Wheels 1154
Wheel-slip control systems, chassis
control and active safety 1260
Wheel-speed sensor 1742
Wheel-speed sensors, wheel-slip
control systems 1265, 1270
Wheel-torque distribution
in the lateral direction 1306
White malleable cast iron,
metallic materials 277
Whitworth pipe threads,
threaded fasteners 446
Wide-band antennas,
electrical engineering 129
Wide-band mufflers,
air-intake system 758
Wide-band λ sensor 1769
Wide-band λ sensor,
exhaust-gas treatment 880
Width requirement, driving dynamics –
commercial vehicles 495
Wind deflector, aerodynamics 522
Wind gas,
engines fueled by natural gas 894

Wind tunnel balance, wind tunnels 525
Wind tunnel variants 527
Windbreak 523
Windingless rotors, alternators 1504
Window bag,
occupant-protection systems 1421
Window drive,
power-window systems 1404
Window panes,
automotive glazing 1374
Windshield and rear-window washer systems, windshield, rear-window and headlamp cleaning systems 1389
Windshield, rear-window and headlamp cleaning systems,
vehicle bodies 1380
Windshields with sunshield coating,
automotive glazing 1376
Winglet, EGR cooler 744
Winter tires 1186
Winter tires, tire tread 1174
Winter-tire designation 1184
Wiper arms, windshield, rear-window and headlamp cleaning systems 1385
Wiper blades, windshield,
rear-window and headlamp cleaning systems 1386
Wiper linkage, windshield,
rear-window and headlamp cleaning systems 1384
Wiper potentiometer,
angular-position sensors 1724
Wiper-blade element,
windshield, rear-window and headlamp cleaning systems 1387
Wipers,
rod and piston sealing systems 411
Wired-and arbitration scheme,
CAN 1694
Wire-fabric jacket,
exhaust-gas system 796
Wireless signal transmission,
radio and TV broadcasting 1794
Wiring harnesses 1518
WLTC, test cycles 1060
WLTP test procedure, test cycles 1060
Wöhler (fatigue) tests,
material parameters 261
Wöhler curve, material parameters 261
Woodruff-key couplings,
positive or form-closed joints 426
Work diagram, shock absorbers 1141
Work, electrical engineering 112
Work, mechanics 47
Work, thermodynamics 87
Worked penetration, lubricants 543
Working cylinder, steering 1197
Workshop diagnostic modules,
ECU diagnostics 1106
Workshop diagnostics,
ECU diagnostics 1106
World harmonized cycles, test cycles for heavy commercial vehicles 1070
World Harmonized Not-to-Exceed Zone, test cycles for heavy commercial vehicles 1065
World Harmonized Stationary Cycle, test cycles for heavy commercial vehicles 1070
World Harmonized Transient Cycle, test cycles for heavy commercial vehicles 1070
Worldwide Light-duty Test Cycles,
test cycles 1060
Wrap angle, belt drives 721
Wrap belt, step-up gear unit 622
Wrap chain, step-up gear unit 622
Wrong-way driver warning, information and warning systems 1904
Wrought alloys, nonferrous metals 277

X

X by wire, mechatronics 1640
X split, antilock braking system 1267
X-capacitors, inverters 1585
Xenon headlamps,
lighting equipment 1348
Xenon light, lighting equipment 1351
X-ray diffraction, elements 188
X-ray fluorescence analysis,
elements 187
X-ray photoelectron spectroscopy,
elements 188

Y

Yaw angle, vehicle motion 463
Yaw gain, motor-vehicle dynamics 481
Yaw moment, vehicle motion 463
Yaw rate 1743
Yaw, vehicle motion 463
Yaw-moment build-up delay,
antilock braking system 1262
Y-capacitors, inverters 1585
Yield point, lubricants 539
Yield point, material parameters 260
Yield point, mechanics 51
Yield strength, material parameters 260

Z

Zener breakdown, diode 140
Zener diode, electronics 142
Zero delivery calibration, common rail 928
Zero frame, Serial Wire Ring 1712
Zero points, mathematics 205
Zero-ampere speed, alternators 1498
Zero-order reaction, reaction kinetics 196
Zeroth law, thermodynamics 88
Zero-voltage vector, inverters 1587
ZEV program, CARB emission-control legislation 1046
Zinc deposits, Deposits and coatings 348
Zinc-lamella deposits, Deposits and coatings 350
Zinc-phosphate coatings, Deposits and coatings 353
Zirconium-oxide ceramic, chemistry 203
Zone ECUs, architecture of electronic systems 1670

Abbreviations

Symbols

3PMSF Three Peak Mountain Snow Flake (snow-flake symbol)

A

AAS Atomic Absorption Spectroscopy
ABS Acrylnitrile-Butadien-Styrene
ABS Antilock Braking System
ABV Antilock braking system (German: Antiblockiervorrichtung)
AC Accessory Control
AC Alternating Current
ACC Adaptive Cruise Control
ACEA Association des Constructeurs Européens de l'Automobile (Association of European Automobile Manufacturers)
ACT Advanced Clean Truck Regulation
ADAC Allgemeiner Deutscher Automobil-Club
ADAS Advanced Driver Assistance Systems
ADASIS Advanced Driver Assistance Systems Interface Specifications
ADB Adaptive Driving Beam
ADC Analog-Digital Converter
ADR Accord européen relatif au transport international des marchandises Dangereuses par Route (European Agreement concerning the International Carriage of Dangerous Goods by Road)
ADR Advanced Digital Receiver
ADR Australian Design Rule
AE Automotive Electronics
AEB Automatic Emergency Braking
AEBS Advanced Emergency Braking System
AF Antiferromagnetic
AF Automated driving
AFC Anti-Friction-Coatin
AFS Adaptive Frontlighting System
AGI Air-Gap-Insulated
AGM Absorbent Glass Mat
AHP Accelerator Heel Point
AI Articulation Index
AI Artificial Intelligence
AKF Carbon canister (German: Aktivkohlefalle)
ALB Automatic load-sensitive braking-force metering (German: Automatische lastabhängige Bremskraftregelung)
ALU Arithmetic-Logic Unit
AM Amplitude Modulation
AMB Active Metal Brazing
AMLCD Active Matrix Liquid Crystal Display
AMR Anisotropic Magneto-Resistance
AMT Automated Manual Transmission
ANC Active Noise Control
ANN Artificial Neural Network
ANSI American National Standards Institute
API American Petroleum Institute
API Application Programming Interface
APS Active Pixel Sensor
APSC All-Pole Short-Circuiting
APSM Advanced Porous Silicon Membrane
ARAS Advanced Rider Assistance Systems
ARD Active Rectification Diode
ARLA Agente Reductor Liquido de Òxido de Nitrogénio Automotivo
AS Air System
ASAM Association for Standardization of Automation and Measuring Systems
ASF Audi Space Frame (aluminum frame structure of supporting integrated sheet panels from Audi)
ASIC Application Specific Integrated Circuit
ASIL Automotive Safety Integrity Level
ASM Asynchronous machine
ASM Trailer control module (German: Anhängersteuermodul)
ASPICE Automotive Software Process Improvement and Capability Determination

ASR Traction control system (German: Antriebsschlupf-regelung)
ASS Anti-snow system
ASTM American Society of Testing and Materials
ASU Automatic interference suppression (German: Automatische Störunterdrückung)
ASVP Air Saturated Vapor Pressure
ASW Application software
AT Advanced Technology
AT Automated Transmission
ATCT Ambient Temperature Correction Test
ATF Automatic-Transmission Fluid
ATL Exhaust-gas turbochargers (German: Abgasturbolader)
AUS Aqueous Urea Solution
AUTOSAR Automotive Open Systems Architecture

B

BBA Service-Brake System (German: Betriebsbrems-anlage)
BCD Bipolar, CMOS, DMOS
BCI Bulk Current Injection
BCM Body Computer Module
BCU Battery Control Unit
BDC Bottom Dead Center (piston of internal-combustion engine)
BDE Gasoline Direct Injection (German: Benzin-Direktein-spritzung)
BEM Boundary-Element Method
BETP Bleed Emissions Test Procedure
BEV Battery Electric Vehicle
BIP Beginning of Injection Period
BIPM Bureau International des Poids et Mesures (International Bureau of Weights and Measures)
BL Base Line
BLDC Brushless Direct Current
BLE Bluetooth Low Energy
BM Bus-Minus
BMD Bag Mini Diluter
BMS Battery Management System
BP Bus-Plus
BPP Bipolar Plate
BRM Boost Recuperation Machine
BRS Boost Recuperation System
BSS Byte-Start-Sequence
BSW Basic Software
BtL Biomass-to-Liquid
BWC Butane Working Capacity
BZE Battery-Status Recognition (German: Batteriezustands-erkennung)

C

C2C Car-to-Car Communication
C2I Car-to-Infrastructure Communication
CAD Computer-Aided Design
CAE Computer-Aided Engineering
CAFC Corporate Average Fuel Consumption
CAFE Corporate Average Fuel Economy
CAL Computer-Aided Lighting
CAN Controller Area Network
CAP Compliance Assurance Program
CARB California Air Resource Board
CBS Combined Brake System
CCC Car Connectivity Consortium
CCD Charge-Coupled Device
CCD Complex Device Drivers
CCFL Cold-Cathode Fluorescence Lamp
CCL Connected-Component Labeling
CCM Continuous Conduction Mode
CCMC Comité des Constructeurs d'automobiles du Marché Commun (CCMC manufacturers committee)
CD Charge-Depleting Test
CD Compact Disc
CDPF Catalyzed Diesel Particulate Filter
CDR Clock Data Recovery
CE Conformité Européenne
CE Coordination Engine
CEMEP European Committee of Manufacturers of Electrical Machines and Power Electronics
CEN Comité Européen de Normalisation (European Committee for Standardization)
CF Carbon Fibers

CF Conformity Factor
CFD Computational Fluid Dynamics
CFPP Cold Filter Plugging Point (filterability limit)
CFR Code of Federal Regulations of the United States
CFR Cooperative Fuel Research
CFV Critical Flow Venturi
CGI Compacted Graphite Iron
CGPM Conférence Générale des Poids et Mésures (General Conference on Weights and Measures)
CGW Central Gateway
CHD Case Hardening Depth
CI Cetane Index
CIP-GMR Current-in-Plane GMR
CISC Complex Instruction Set Computing
CISPR Comité International Spécial des Perturbations Radioélectriques (International Special Committee for Radio Interference)
CL Classic Line
CLD Chemiluminescence Detector
CMMI Capability Maturity Model Integration
CMOS Complementary Metal Oxide Semiconductor
CMRR Common Mode Rejection Ratio
CMVSS Canadian Motor Vehicle Safety Standard
CN Cetane Number
CNG Compressed Natural Gas
CNN Convolutional Neural Network
CO Communication
COD Conversion-of-Olefins-to-Distillates
CoM Change of Mind
COM Communication
COP Child Occupant Protection
COP Conformity of Production
CP Circular Pitch
CPC Condensation Particulate Counter
CPI Compression Pre-Ignition number
CPP-GMR Current Perpendicular-to-Plane GMR
Cpsi Channels per Square Inch
CPU Central Processing Unit
CR Chloroprene Rubber
CR Common Rail
CRC Cyclic Redundancy Check
CRS Common-Rail-System
CRT Continuously Regenerating Trap
CS Charge-Sustaining Test
CS Checksum
cs cold stabilized
CSC Cell Supervisory Circuit
CSM Chlorosulfonated polyethylene
CST Controllo Stabilità e Trazione
ct cold transient
CtL Coal-to-Liquid
CTS Conti Tire System
CV Computer vision
CVD Chemical Vapor Deposition
CVS Constant Volume Sampling
CVT Continuously Variable Transmission
CW Continuous Wave
CWS Collision Warning System
C-WTVC China World Transient Vehicle Cycle

D

DAB Digital Audio Broadcasting
DAC Digital-Analog Converter
DAPRA Defense Advanced Research Projects Agency
DC Diffusion Charger
DC Direct Current
DCB Direct Copper-Bonding
DCN Derived Cetane Number
DCT Dual Clutch Transmission
DCT Dynamic Coupling Torque
DCT-C Dynamic Coupling Torque at Center
DDA Digital Directional Antenna
DDS Digital Diversity System
DEF Diesel Exhaust Fluid
DEQ Digital Equalizer
DF Dynamo Field
DFC Data Frame Counter
DFM Dynamo-Field Monitor
DFV Vapor/Liquid Ratio (German: Dampf-Flüssigkeits-Verhältnis)
DGDI-S Diesel-Gasoline-Direct-Injection Solenoid
DGL Differential Equation (German: Differentialgleichung)
DGPS Differential GPS
DHT Dedicated Hybrid Transmission
DI Direct Injection
DIN German Institute for Standardization
DIS Draft International Standard

DK Dielectric Constant (German: Dielektrizitätskonstante)
DLC Diamond Like Carbon
DLC Double Layer Capacitor
DLP Digital Light Projection
DMA Direct Memory Access
DME Dimethyl Ethers
DMIPS Dhrystone Million Instructions Per Second
DMOS Double-Diffused Metal-Oxide Semiconductor
DMS Differential Mobility Spectrometer
DMS Strain Gages (Strain-Gage Resistor) (German: Dehnmess-streifen)
DNC Dynamic Noise Covering
DNN Deep Neural Network
DOC Diesel Oxidation Catalyst
DOE Design of Experiment
DOHC Double Overhead Camshaft
DOT Department of Transportation (USA)
DP Dynamic Programming
DPF Diesel Particulate Filter
DPMO Defects per Million Opportunities
DRAM Dynamic Random-Access Memory
DR-F Flat-seat pressure-control valve
DRIE Deep Reactive Ion Etching
DRM Digital Radio Mondial
DRO Dielectric Resonance Oscillator
DR-S Spool pressure-control valve
DS Diagnostic System
DSA Digital Sound Adjustment
DSC Dynamic Stability Control
DSM Diagnostic System Management
DSP Digital Signal Processor
DST Direct-Shift Transmission
DSTC Dynamic Stability and Traction Control
DSTN-LCD Double layer Super Twisted Nematic Liquid Crystal Display
DTC Diagnostic Trouble Code
DTM Deutsche Tourenwagen Masters
DTS Dynamic-Trailing-Sequence
DVB Digital Video Broadcasting
DVB-C Digital Video Broadcasting Cable
DVB-H Digital Video Broadcasting Handhelds
DVB-S Digital Video Broadcasting Satellite
DVB-T Digital Video Broadcasting Terrestrial
DVD Digital Versatile Disc
DVPE Dry Vapor Pressure Equivalent
DVR Compression set
DWT Dynamic Wheel Torque Distribution
E/E Electrics/Electronics

E

EAB Electronic shutdown (German: Elektronische Abstellung) (diesel control)
EAC Electronic Air Control
EB Electrical steel sheet and strip (German: Elektroblech und -band)
EBA Emergency Brake Assist
EBD Electronic braking-force distribution
EBD Electronic Brakeforce Distribution
EBS Electronic Battery Sensor
EBS Electronic Braking System
EC Electronic Commutated
EC European Community
EC Exhaust Closes
eCall emergency Call
eCBS electronically Combined Brake System
ECC Error Correction Code
ECE Economic Commission for Europe
ECMS Equivalent Consumption Minimization Strategy
ECU Electronic Control Unit
ECVT Electrical Continuously Variable Transmission
EDC Electronic Diesel Control
EDR Event Data Recorder
EEA European Economic Area
EEC European Economic Community
EEM Electric Energy Management (vehicle electrical system)
EEPROM Electrically Erasable Programmable Read Only Memory
EEV Enhanced Environmentally-Friendly Vehicle
EFB Enhanced Flooded Battery

EFF Designations of Efficiency classes for electrical machines (German: Bezeichnung von Effizienzklassen für elektrische Maschinen)
EFM Exhaust Flow Meter
EFU Activation Spark Suppression (German: Einschaltfunken-unterdrückung)
EG European Community (German: Europäische Gemeinschaft)
EG Self-Steering Gradient (German: Eigenlenkgradient)
EGR Exhaust-Gas Recirculation
EGS Electronic Transmission Control (German: Elektronische Getriebesteuerung)
EH Extended Hump
EHB Electrohydraulic Brake
EHC Hydrated Ethanol Fuel
EHD Elastohydrodynamics
EHF Extremely High Frequency
EIPR Equivalent Isotropic Radiated Power
EIR Emissions Information Report
EIS Electrochemical Impedance Spectroscopy
EIW Equivalent Inertia Weight
EKP Electric Fuel Pump (German: Elektrokraftstoffpumpe)
EL Efficiency Line
ELPI Electrical Low Pressure Impactor
ELR European Load Response
EM Electrical Machine
EM Monomode Fiber (German: Einmodenfaser)
EMC Electromagnetic Compatibility
EMD Engine Manufacturer Diagnostics
EMP Electro-Mechanical Parking brake
EN European Standard
EO Exhaust Opens
EOBD European On-Board Diagnostics
EOL End of Life
EOL End of Line
EP Epoxy
EP Extreme Pressure
EPA Environment Protection Agency
EPB Electronic Parking Brake
EPDM Ethylene Propylene Diene Monomer
EPM Engine & Powertrain Manager
EPROM Erasable Programmable Read Only Memory
ERBP Equilibrium Reflux Boiling Point
ES Exhaust System
ESC Electronic Stability Control
ESC Energy Supply Company
ESC European Steady-State Cycle
ESD Electrostatic Discharge
ESI Electronic Service Information
ESM Electrically Excited Synchronous Machine
ESP® Electronic Stability Program
ESR Enhanced Seat Belt Reminder
ESS Energy-Saving System
ET Rim Offset (German: Einpresstiefe)
ETBE Ethyl Tertiary Butyl Ether
ETC Electronic Throttle Control
ETC European Transient Cycle
ETFE Ethylene-Tetrafluoroethylene
ETK Emulator Probe (German: Emulator-Tastkopf)
ETN European Type Number
ETRTO European Tyre and Rim Technical Organisation
EU European Union
euATL Electrically Assisted Exhaust-Gas Turbocharger (German: Elektrisch unterstützter Abgasturbolader)
EUC Enhanced Understeering Control
EUDC Extra Urban Driving Cycle
EUV Electric Changeover Valve (German: Elektroumschaltventil)
EV Electric Vehicle
EV Exhaust Valve
EV Fuel Injector (German: Einspritzventil)
EVAP Evaporation
EWIR Emissions Warranty Information Report

F

FAEE Fatty acid ethyl ester
FAME Fatty Acid Methyl Ester
FAS Driver-Assistance System (German: Fahrerassistenzsystem)

FBA Parking-Brake System (German: Feststellbremsanlage)
FC Fuel Cell
FCHV Fuel Cell Hybrid Vehicle
FC-REX Fuel Cell Range Extender
FEA Finite-Element Analysis
FEM Finite-Element Method
FEP Fluorinated Ethylene Propylene
FES Frame-End-Sequence
FET Field-Effect Transistor
FFV Flexible Fuel Vehicle
FGR Vehicle-Speed Controller (German: Fahrgeschwindigkeitsregler)
FH Flat Hump
FID Flame Ionization Detector
FiL Function in the Loop
FIR Far Infrared
FIR Field Information Report
FIS Driver-Information System (German: Fahrerinformationssystem)
FKM Fluorocarbon Rubber (German: Fluorkarbon-Kautschuk)
FKM Forschungskuratorium Maschinenbau
FL Free Layer
FLO Fast-Light-Off
FM Frequency Modulation
FMCW Frequency Modulated Continuous Wave
FMEA Failure Mode and Effects Analysis
FMVSS Federal Motor Vehicle Safety Standard (USA)
FoV Field of View
FPA Focal Plane Array
FPK Freely Programmable Instrument Cluster (German: Frei programmierbares Kombiinstrument)
fps frames per second
FRAM Ferroelectric Random-Access Memory
FS Fuel System
FSN Filter Smoke Number
FSR Full Speed Range
FSS Frame-Start-Sequence
FST Federal Smoke Test
FTA Fault Tree Analysis
FTIR Fourier Transform Infrared spectroscopy
FTP Federal Test Procedure
FuSi Functional safety
FW Front Wheel

G

GC Gas-Chromatography Column
GCC Global Chassis Control
GCU Glow Control Unit
GDL Gas-Diffusion Layer
GF Glass Fibers
GIDAS German In-Depth Accident Study
GJL Gray Cast Iron with Lamellar Graphite
GJS Gray Cast Iron with Spheroid Graphite
GKZ Spheroidizing
GLP Glow Plug
GMA Yaw-Moment Build-Up Delay (German: Giermomentaufbauverzögerung)
GMR Giant Magneto-Resistance
GPRS General Packet Radio Service
GPS Global Positioning System
GSM Global System for Mobile Communication
GSR General Safety Regulation
GSY Speed-rating symbol (tires) (German: Geschwindigkeitssymbol)
GTDC Gas-Exchange TDC (Top Dead Center in the exhaust cycle)
GtL Gas-to-Liquid
GTM General Timer Module
GTR Global Technical Regulation
GVW Gross Vehicle Weight
GWP Global Warming Potential
GWR Gross Weight Rating
GZS Glow Control Unit (German: Glühzeitsteuergerät)

H

HB Brinell Hardness
HBA Secondary-Brake System (German: Hilfsbremsanlage)
HC Hydrocarbon
HCCI Homogeneous Charge Compression Ignition
HCI Hydro Carbon Injection
HD Heavy Duty

HDDTC Heavy-Duty Diesel Transient Cycle
HDEV High-Pressure Fuel Injector (German: Hochdruck-Einspritzventil)
HDG Light/Dark Boundary (German: Hell-Dunkel-Grenze)
HDK Half-Differential Short-Circuiting-Ring Sensor (German: Halbdifferential-Kurzschlussringsensor)
HDP High-Pressure Pump (German: Hochdruckpumpe)
HDPE High-Density Polyethylene
HDR High Dynamic Range
HDS Heavy Duty Second Generation
HDV Heavy-Duty Vehicle
HED High-Efficiency Diodes
HEV Hybrid Electric Vehicle
HF High Frequency
HFM Hot-Film Air-Mass Meters
HFRR High-Frequency Reciprocating Rig
HGV Hydrogen Gas Vehicle
HiL Hardware in the Loop
HLDT Heavy Light-Duty Truck
HLM Hot-Wire Air-Mass Meters
HMI Human Machine Interface
HMM Homogeneous Lean
HNBR Hydrogenated Nitrile Butadiene Rubber
HNS Homogeneous Numerically Calculated Surface
HOG Histogram of Oriented Gradients
hp horsepower
HRC Rockwell Hardness, scale C
hs hot stabilized
HSLA High Strength Low Alloy
HSV High-Pressure Switching Valve
ht hot transient
HTD High Torque Drive
HTHS High Temperature High Shear
HUD Head-Up Display
HV High Voltage
HVA Hydraulic Valve-Clearance Compensation
HVO Hydro-treated Vegetable Oil
HWL Urea/Water Solution
I/M Inspection and Maintenance

I

IC Intake Closes
IC Integrated Circuit
ICA Integrated Cruise Assist
ICC Integrated Chassis Control
ICE Intercity Express
ICE Internal-Combustion Engine
ICM Integrated Chassis Management
ICN Indicated Cetane Number
ICP-OES Inductively Coupled Plasma Optical Emission Spectrometry
ID Identifier
IDF Integrated Collision Detection Front
IDI Indirect Injection
IEC International Electrotechnical Commission
IEEE Institute of Electrical and Electronics Engineers
IF Intermediate Frequency
IGBT Insulated Gate Bipolar Transistor
IGES Initial Graphics Exchange Specification
IHU Integrated Head Unit
IHU Internal High-Pressure Forming (German: Innenhochdruckumformung)
IIR Isobutene-Isoprene Rubber
ILSAC International Lubricants Standardization and Approval Committee
IMA Injection-Quantity Compensation (German: Injektormengenabgleich)
IMC Integrated Magnetic Concentrator
IMG Integrated Motor-Generator
IMM Interacting Multiple Model
IMR Integrated Mechanical Relay
IO Intake Opens
IP Ignition Point
IP Internet Protocol
IPB Integrated Power Brake
IPO Input – Processing – Output
IR Individual Control (ABS) (German: Individualregelung)
IR Infrared
IRHD International Rubber Hardness Degree
IRM Individual Control, Modified (ABS) (German: Individualregelung modifiziert)

IS Ignition System
ISC In-Service Conformity check
ISM Industrial, Scientific and Medical
ISO International Organization for Standardization
IT Information Technology
IT Isolé Terre
ITDC Ignition TDC (Top Dead Center in the power cycle)
ITO Indium Tin Oxide
ITSEC Information Technology Security Evaluation Criteria
ITU International Telecommunications Union
IUC In Use Compliance
IUMPR In-Use Monitor Performance Ratio
IUPAC International Union of Pure and Applied Chemistry
IUPR In-Use Performance Ratio
IV Intake Valve

J

JAMA Japan Automobile Manufacturers Association
JASPAR Japan Automotive Software Platform Architecture
JFET Junction-gate Field-Effect Transistor

K

KAMA Korea Automobile Manufacturers Association
KBA Federal Road Transport Office (German: Kraftfahrt-Bundesamt)
Kfz Motor Vehicle (German: Kraftfahrzeug)
KOM Motor Bus (German: Kraftomnibus)
KP Critical Point (German: Kritischer Punkt)
KTL Cathodic Deposition (German: Kathodische Tauchlackierung)
KTS Diagnostics Hardware Small Testers (German: Diagnostics Hardware Kleintester)
KW Crankshaft (German: Kurbelwelle)

L

LAC Load Adaptive Control
LAN Local Area Network
lb pound
lbs pounds
LCA Life Cycle Analysis
LCD Liquid Crystal Display
LCV Light Commercial Vehicles
LDR Boost pressure
LDR Light-Dependent Resistor
LDT Light-Duty Truck
LDV Light-Duty Vehicle
LDW Lane Departure Warning
LDWS Lane Departure Warning System
LED Light-Emitting Diode
LEV Light Electric Vehicle
LEV Low-Emission Vehicle
LF Low Frequency
LGS Linear Equation System (German: Lineares Gleichungssystem)
LI Load Index (tires)
Lidar Light Detection and Ranging
LIF Alternator Liquid Cooling (German: Lichtmaschine Flüssigkeitskühlung)
LIN Local Interconnect Network
LKS Lane Keeping Support
Lkw Truck (German: Lastkraftwagen)
LL Lower Limit
LLDT Light Light-Duty Truck
LLK Intercooling (charge-air cooling) (German: Ladeluftkühlung)
LLR Smooth-Running Control (German: Laufruheregelung)
LMM Air-flow Sensor (German: Luftmengenmesser)
LNG Liquefied Natural Gas. Liquid Natural Gas
LP Low Power
LPG Liquefied Petroleum Gas
LPTC Low Powered Vehicle Test Cycles
LR Load Response
LRF Load Response Driving (German: Load Response Fahrt)
LRR Long-Range Radar
LRS Load Response Start
LSF Low Speed Following
LSG Laminated Safety Glass
LSI Large Scale Integration
LSTM Long Short-Time Memory

LSU Lambda Oxygen Sensor (universal or wideband sensor)
LT Low Temperature
LTE Long Term Evolution
LV Low Voltage
LVDS Low Voltage Differential Signing
LVW Loaded Vehicle Weight (vehicle tare weight plus 300 lb)
LW Long Wave
LWL Optical Fibers (German: Lichtwellenleiter)
LWS Steering-Angle Sensor (German: Lenkwinkelsensor)

M

M+S Mud and Snow
MAC Message Authentication Code
MAC Multiply Accumulate
MAG Active-Gas Metal-Arc (German: Metall-Aktivgas)
MAMAC MOST Asynchronous Medium Access Control
MB Mercedes Benz
MCAL Micro-Controller Abstraction Layer
MCD Measurement, Calibration and Diagnosis
MCU Micro Controller Unit
MDPV Medium-Duty Passenger Vehicle
MDV Medium-Duty Vehicle
MEMS Micro-Electro-Mechanical Systems
MF Medium Frequency
MH Metal Hydride
MIG Metal Inert Gas
MIL Malfunction Indicator Lamp
MIL Military
MiL Model in the Loop
MIM Metal Injection Molding
MIMO Multiple Input Multiple Output
MISRA Motor Industry Software Reliability Association
MKS Multi-Body Simulation (German: Mehrkörpersimulation)
ML Maturity Level
MM Multi-Mode Fiber
MMA Mean-Quantity Adaptation (German: Mengenmittelwert-adaption)
MMIC Millimeter-Wave Integrated Circuit
MMT Methylcyclopentadienyl Manganese Tricarbonyl
MNEDC Modified New European Driving Cycle
MO Monitoring
MON Motor Octane Number
MOS Metal-Oxide Semiconductor
MOSFET MOS Field-Effect Transistor
MOST Media-Oriented Systems Transport
MP Momentary Pole
MPC Model Predictive Control
MPEG Moving Picture Experts Group
mpg Miles Per Gallon
MPP Most Probable Path
MPT Multi-Purpose Tire
MPU Micro Processor Unit
MRAM Magnetoresistive Random-Access Memory
MRR Mid-Range Radar
MSB Motorized Seat Belt (electro mechanical belt pretensioner)
MSC Micro Second Channel
MSC Motorcycle Stability Control
MSE Motronic Small Engines
MSG Gas-Shielded Metal-Arc (German: Metall-Schutzgas)
MSI Medium-Scale Integration
MSI Multi-Spark Ignition
MSR Engine Drag-Torque Control (German: Motorschlepp-momentregelung)
MT Manual Transmission
MTBE Methyl Tertiary Butyl Ether
MTC Motorcycle Traction Control
MTF Modulation Transfer Function
MtG Methanol-to-Gasoline
MTTF Mean Time To Failure
MW Medium Wave

N

NAFTA North American Free Trade Agreement
NAO Non Asbestos Organics
NBR Nitrile Butadiene Rubber
NCAP New Car Assessment Program
NCS Needle Closing Sensor
NDIR Non-Dispersive Infrared (analyzer)
NDUV Non-Dispersive Ultraviolet (analyzer)
NEDC New European Driving Cycle
NEV New Energy Vehicle

NF Non-Ferrous
NFC Near Field Communication
Nfz Commercial Vehicle (German: Nutzfahrzeug)
NHD Nitriding Hardness Depth
Nht Nitriding Depth (German: Nitrierhärtetiefe)
NHTSA National Highway Traffic Safety Administration
NIC Network Interface-Controller
NiMH Nickel-Metal-Hydride
NIR Near-Infrared
NIT Network Idle Time
Nkw Commercial Vehicles (German: Nutzkraftwagen)
NLGI National Lubricating Grease Institute
NLU Natural Language Understanding
NMHC Non-Methane Hydrocarbon
NML Nonmagnetic Layer
NMOG Non-Methane Organic Gases
NMOS N-channel MOS Transistor
NN Neural Network
NR Natural Rubber
NRZ Non Return to Zero
NSC NO_x Storage Catalyst
NSM Shunt-Wound Machine (German: Nebenschlussmaschine)
NTC Negative Temperature Coefficient
NTE Not To Exceed
NVH Noise Vibration Harshness
NYCC New York City Cycle

O

OAT Organic Acid Technology
OBD On-Board Diagnostics
OC Occupant Classification
OCV Open Circuit Voltage
OD Operating Data
ODB Offset Deformable Barrier Crash
ODD Operational Design Domain
OEM Original Equipment Manufacturer
OHC Overhead Camshaft
OHV Overhead Valves
OLED Organic Light Emitting Diode
OME Oxymethylene Ethers
OMM Surface Micromechanics (German: Oberflächenmikromechanik)
OP Operational Amplifier (German: Operationsverstärker)
OPNV Local Public Transport (German: Öffentlicher Personennahverkehr)
OPV Operational Amplifier (German: Operationsverstärker)
OS Operating System
OSEK Open Systems and their Interfaces for Electronics in Motor Vehicles
OSI Open Systems Interconnection
OTX Open Test Sequence Exchange Format
Overlap TDC Overlap TDC (Top Dead Center in the exhaust cycle)

P

PA Polyamide
PAO Poly-α-olefin
PAR Parity
PAS Peripheral Acceleration Sensor
PAX Pneu Accrochage, X = synonym for Michelin radial tire technology
PBT Polybutylene Terephthalate
PC Passenger Car
PC Personal Computer
PC Polycarbonate
PCB Printed Circuit Board
PCM Phase Change Memory
PCP Pre-Crash Positioning
PCV Positive Crankcase Ventilation
PCW Predictive Collision Warning
PDP Positive Displacement Pump
PDU Protocol Data Unit
PE Polyethylene
PEBS Predictive Emergency Braking System
PEEK Polyetheretherketone
PEM Polymer Electrolyte Membrane
PEM Proton Exchange Membrane
PEMS Portable Emission Measurement System
PES Poly-Ellipsoid System (headlamp)
PET Polyethylene Terephthalate
PF Phenol Formaldehyde
PFA Perfluoroalkoxy (polymers)
PFC Power Factor Correction

PGM Probabilistic Graphical Model
pH Potentia Hydrogenii
PHEV Plug-in Hybrid Electric Vehicle
PHV Plug-in Hybrid Electric Vehicle
PK Profile Code (German: Profilkurzzeichen)
Pkw Passenger Car (German: Personenkraftwagen)
PL Pinned Layer
PL Power Density Line
PLL Phase-Locked Loop
PM Particulate Matter
PM Proportional solenoid
PMD Paramagnetic Detector
PMD Photonic Mixing Device
PMM Polymethylmethacrylate
PMMA Polymethylmethacrylate
PMOS P-channel MOS Transistor
PN Particle Number
PNLT Post New Long Term
POF Polymer Optical Fibers
POI Points of Interest
POM Polyoximethylene
PP Polypropylene
PPA Polyphthalamide
PPNLT Post Post New Long Term
PPS Peripheral Pressure Sensor
PPS Polyphenylene Sulfide
PROM Programmable Read Only Memory
PS Polystyrene
PSD Power Spectral Density
PSFB Phase Shifted Full Bridge
PSI Peripheral Sensor Interface
PSM Passive Safety Manager
PSM Permanent-field Synchronous Machine
PSOC Partial State of Charge
PS-P Presafe Pulse
PSRR Power-Supply Rejection Ratio
PTB Physikalisch-Technische Bundesanstalt (National Metrology Institute of Germany)
PTC Positive Temperature Coefficient
PTFE Polytetrafluoroethylene (Teflon)
PtG Power-to-Gas
PtL Power-to-Liquid
PU Polyurethane
PUR Polyurethane
PVB Polyvinyl Butyral
PVC Polyvinyl Chloride
PVD Physical Vapor Deposition
PVDF Polyvinylidene Fluoride
PVE Production Vehicle Evaluation
PWM Pulse-Width Modulation
PWR Pulse-Width-Modulation inverter
PZEV Partial Zero-Emission Vehicle
PZT Plumbum Zirconate Titanate

Q

QFD Quality Function Deployment
QM Quality Management
QVGA Quarter Video Graphics Array

R

RA Rear Axle
Radar Radio Detection and Ranging
RAM Random Access Memory
RAMSIS Computer-aided anthropological-mathematical system for passenger simulation (German: Rechnerunterstütztes antropometrisch-mathematisches System zur Insassensimulation)
RBU Redundant Brake Unit
RDE Real Driving Emissions
RDKS Tire-Pressure Monitoring System (German: Reifendruckkontrollsystem)
RDS Radio Data System
REX Range Extender
RF Radio Frequency
RFA X-ray Fluorescence Analysis (German: Röntgenfluoreszenzanalyse)
RFCMOS Radio-Frequency CMOS
RGM Reliability-Growth Management
RIM Reaction-Injection Molding
RISC Reduced Instruction-Set Computing
RME Rapeseed Methyl Ester (alternative fuel)
RMF Rollover Mitigation Functions
RMQ Root Mean Quad
RMS Root Mean Square
RMV Rollover Mitigation Function
ROM Read Only Memory
RON Research Octane Number
ROV Rotating High-Voltage Distribution (German: Rotierende Hochspannungsverteilung)

RR Rolling Resistance
RREG Council Directive of the European Community (today: European Union)
RRIM Reinforced Reaction-Injection Molding
RSM Series-Wound Machine (German: Reihenschluss-maschine)
RTA Real Time Architect
RTE Run-Time Environment
RUV Stationary Voltage Distribution (German: Ruhende Span-nungsverteilung)
RW Rear Wheel
RWAL Rear Wheel Anti Lock Brake System

S

S Center of Gravity
SAE Society of Automotive Engineers
SAF Synthetic Antiferromagnetic Free Layer
SBC Sensotronic Brake Control
SBI Suppression of Background Illumination
SBR Seat Belt Reminder
SBR Styrene Butadiene Rubber
SC System Control
SCM Secondary Collision Mitigation
SCR Selective Catalytic Reduction
SCU Sensor Control Unit
SD Secure Digital
SD Sigma-Delta
SD System Document
SDARS Satellite Digital Audio Radio Service
SEI Software Engineering Institute
SEL Sound Exposure Level
SENT Single Edge Nibble Transmission
SEW Steel and Iron Material Specification Sheets (German: Stahl-Eisen-Werkstoffblätter)
SfM Structure of Motion
SFTP Supplemental Federal Test Procedure
SG Float-Angle Gradient
SH Sample and Hold
SHD Surface Hardness Depth
SHED Sealed Housing for Evaporative Determination
SHF Super High Frequency
SI Speed Index, speed symbol (tires)
SI Système International (International System of Units)
SIL Safety Integrity Level
SiL Software in the Loop
SIR Styrene-Isoprene Rubber
SIS Service Information System
SL Select-Low (ABS)
SLADR Slave Address
SLC Sensor-Less Control
SLI Starting, Lighting, Ignition
SM Synchronous Machine
SMD Surface-Mounted Device
SMG Separate Motor-Generator
SMK Inertia Weight Class (German: Schwungmassenklasse)
SML Side-Marker Lamps
SMPS Scanning Mobility Particle Sizer
SMS Short Message Service
SMT Surface-Mount Technology
SNR Signal-to-Noise Ratio
SOC State of Charge (battery)
SoC System on Chip
SOF State of Function (battery performance)
SOH State of Health (battery)
SOTIF Safety of the Intended Function
SPAD Single Photon Avalanche Diode
SPC Statistical Process Control
SPI Serial Peripheral Interface
SPICE Software Process Improvement and Capability Determination
SR Slew Rate
SRAM Static Random-Access Memory
SRET Scanning Reference-Electrode Techniques
SRM Switched Reluctance Machine
SRR Short-Range Radar
SSI Small-Scale Integration
SSV Side-by-Side Vehicle
STEP Standard for the Exchange of Product Model Data
STN LCD Super Twisted Nematic Liquid Crystal Display
StVZO German road traffic licensing regulations (German: Straßen-verkehrs-Zulassungsordnung)
SULEV Super Ultra-Low-Emission Vehicle

SURF Speeded Up Robust Features
SUV Sport Utility Vehicle
SV Switchover Valve
SVM Space Vector Modulation
SVPWM Space Vector Pulse Width Modulation
SW Short Wave
SW Software
SWC Software Component
SWR Serial Wire Ring

T

TA Type Approval
TC Transient Cycle
TCM Trailer Control Module
TCO Total Cost of Ownership
TCP Transmission Control Protocol
TCU Transmission Control Unit
TD Torque Demand
TDC Time to Digital Converter
TDC Top Dead Center (piston of internal-combustion engine)
TDMA Time Division Multiplex Access
TEM Transversal Electromagnetic Mode
TFT Thin-Film Transistor
THC Total Hydrocarbons
TKU Technical customer documents (German: Technische Kunden-unterlagen)
TLEV Transitional Low-Emission Vehicle
TMC Traffic Message Channel
TMR Tunnel-Magnetoresistive Effect
TN LCD Twisted Nematic Liquid Crystal Display
TOF Time of Flight
TP Triple Point
TPA Thermoplastic Copolyamides
TPC Thermoplastic Copolyester Elastomers
TPE Thermoplastic Elastomers
TPEG Transport Protocol Experts Group
TPMS Tire Pressure Monitoring System
TPO Thermoplastic Elastomers on an Olefin basis
TPU Thermoplastic Elastomers on a Urethane Basis
TPV Cross-linked thermoplastic elastomers on an olefin basis
TRX Tension Répartie (distributed tension)
TS Technical Specification
TS Torque Structure
TSG Toughened Safety Glass
TSM Trailer Sway Mitigation
TSS Transmission Start Sequence
TTB Time To Brake
TTC Time To Collision
TTCAN Time Triggered CAN
TTL Transistor-Transistor Logic (bipolar integrated digital circuit)
TTS Time To Steer
TtW Tank to Wheel
TV Television
TWC Three-way Catalytic Converter
TWI Tread Wear Indicator

U

UDC Urban Driving Cycle
UDDS Urban Dynamometer Driving Schedule
UF Utility Factor
UFOME Used Frying Oil Methyl Ester
UHF Ultra-High Frequency
UIS Unit Injector System
UL Upper Limit
ULEV Ultra-Low-Emission Vehicle
ULSD Ultra Low Sulfur Diesel
ULSI Ultra Large Scale Integration
UMTS Universal Mobile Telecommunication System
UN/ECE United Nations/Economic Commission for Europe
UN United Nations
UPS Ultraviolet Photo Spectroscopy
UPS Unit Pump System
USA United States of America
USW Ultra-Short Wave
UTP Unshielded Twisted Pair
UV Ultraviolet
UWB Ultra Wide Band

V

V2G Vehicle to Grid
V2H Vehicle to Home
V2I Vehicle to Infrastructure
V2L Vehicle to Load
V2V Vehicle to Vehicle
V2X Vehicle to Infrastructure
VA Front Axle
VC Vehicle Computer

VCI Volatile Corrosion Inhibitor
VCO Voltage-Controlled Oscillator
VCR Variable Compression Ratio
VCSEL Vertical-Cavity Surface-Emitting Laser
VCU Vehicle Control Unit
VDA German Association of the Automotive Industry (German: Verband der Automobilindustrie)
VDA-FS German Association of the Automotive Industry – Surface Data Interface (German: Verband der Automobilindustrie – Flächenschnittstelle)
VDE Association for Electrical, Electronic and Information Technologies (German: Verband der Elektrotechnik Elektronik Informationstechnik e.V.)
VDI Association of German Engineers (German: Verein Deutscher Ingenieure)
VDM Vehicle Dynamics Management
VDU Vehicle Dynamics Unit
VDX Vehicle Distributed Executive
VE Distributor Injection Pump (German: Verteilereinspritzpumpe)
VECTO Vehicle Energy and Consumption Calculation tool
VFB Virtual Function Bus
VFD Vacuum Fluorescence Design (Vacuum-fluorescence display technology)
VGA Video Graphics Array
VHAD Vehicle Headlamp Aiming Device (adjusters for front headlamps)
VHD Vertical Hall Device
VHF Very High Frequency
VI Viscosity Index
VIS Visible
VLF Very Low Frequency
VLI Vapor Lock Index
VLSI Very Large Scale Integration
VMC Vehicle Motion Control
VMM Volume Micromechanics
VOL Visual Optical Aim Left
VOR Visual Optical Aim Right
VPI Vapor-Phase Inhibitor
VR Distributor Injection Pump, Radial-Piston Pump (German: Verteilereinspritzpumpe, Radialkolbenpumpe)
VRLA Valve-Regulated Lead-Acid
VSA Vehicle Stability Assist
VSC Vehicle Stability Control
VTG Variable Turbine Geometry (exhaust-gas turbocharger)
VVT Variable Valve Timing

W

WdK Economic Association of the German Rubber Industry (German: Wirtschaftsverband der deutschen Kautschukindustrie)
WEC World Endurance Championship
wERBP wet Equilibrium Reflux Boiling Point
WHSC World Harmonized Stationary Cycle
WHTC World Harmonized Transient Cycle
WLAN Wireless Local Area Network
WLTC Worldwide Light-duty Test Cycles
WLTP Worldwide harmonized Light vehicle Test Procedure
WMTC Worldwide harmonized Motorcycle Testing Cycle
WNTE Worldwide harmonized Not To Exceed
WRC World Rally Championship
WtT Well to Tank
WtW Well to Wheel
WWH World Wide Harmonized
WWW World Wide Web

X

XML Extensible Markup Language
XPS X-ray Photoelectron Spectroscopy

Z

ZEV Zero-Emission Vehicle
ZME Metering unit (German: Zumesseinheit)